International Journal of Theoretical Physics, Vol. 42, No. 2, February 2003 (© 2003)

What Is the Lagrangian Counting?

Tommaso Toffoli[1]

Received January 4, 2003

The calculus of variations provides an exhaustive descriptive account of the least-action principle and other icons of nature's parsimoniousness. However, when we seek *explanations* rather than mere formal descriptions of variational principles, almost invariably we discover a *combinatorial* origin; the best known examples are thermodynamics and Darwinian evolution. When it comes to the least-action principle, however, it is surprising that reductionistic attacks of this kind have been virtually absent. An eminent exception is, of course, Richard Feynman's explanation in terms of quantum path integrals, but even there, though the spirit of the approach is combinatorial, the nature of the objects that are "counted" is somewhat elusive; one is still explaining a mystery through another mystery. Feynman himself stresses that "whoever thinks they understand quantum mechanics, they don't," and ultimately admits "I don't know what action is." The challenge we proposes is to devise "classical" models (that is, models based on ordinary counting of large numbers of discrete objects rather than superposition of complex amplitudes) of classical analytical mechanics. Never mind what method nature actually uses; how come models of this kind—which, for example, were a dime a dozen for the laws of perfect gases—are so hard to come by for the least-action principle?

KEY WORDS: least action principle; action as amount of computation; Lagrangian mechanics.

1. INTRODUCTION

No natural action can be abbreviated . . . .
Every natural action is generated by nature in the shortest possible way she can find.
From given causes, effects are generated by nature in the most efficient way.

Leonardo[2]

One may be tempted to construe Leonardo's words (ca. 1500) as the intuition of a lone genius running ahead of the pack and boldly anticipating the variational

[1] Electrical and Computer Engineering, Boston University, Boston, Massachusetts; e-mail: tt@bu.edu.
[2] Leonardo da Vinci: "Nessuna azione naturale si pò abreviare . . . . Ogni azione naturale è generata dalla natura nel più brieve modo che trovar si possa" [Atl. 112 v.a.]; "Data la causa la natura opera l'effetto nel più breve modo che operar si possa" [Ar. 132 r.].

0020-7748/03/0200-0363/0

principles of mechanics. But Leonardo was merely expressing feelings that were widely shared by the "intelligentsia" of the time. The neo-Platonic revival of the Renaissance had a pantheistic conception of Nature and saw her as mystically wise, at one time *providential* and *parsimonious* (cf. Hildebrandt and Tromba, 1996)—in sum, a nature that made use of Operations Research *ante litteram* in husbanding the world.

In any event, Leonardo's words were the harbingers of ideas that gradually took shape through the work of Fermat, Leibniz, Euler, and many others, and, though they continued to evolve after that, were captured in a memorable snapshot by Maupertuis's *law of least action* (1744):

> If there occurs some change in nature, the amount of action necessary for this change must be as small as possible.

Note that by this time *action* had become a technical word, even though some vagueness still remained as to what it referred to. The meaning of "action" was assigned in a final way by Lagrange in his *Méchanique Analitique* (1788), right before the French Revolution. Concurrently, the "law of least action," which for Maupertuis still had a strongly mystical flavor, had matured into the *principle of stationary action*—a coolly detached statement of fact about certain formal aspects of Newtonian Mechanics. As a mathematical physicist, the business of Lagrange was not to prescribe (give teleological motivations) or explain (give reasons why), but to *prove*—that is, extract interesting tautologies.

Yet a principle of such simplicity and power demands an explanation. If, given the laws of Netwonian mechanics, one can conclude that any actual dynamical trajectory is privileged in so far as the action along it is *minimal* (or, at least, stationary) with respect to the swarm of virtual trajectories that surround it, one is certainly entitled to ask, "Yes, but why is Newtonian mechanics such? Why are its trajectories so privileged? Could it not have been otherwise? Is a trajectory taken because it is privileged, or does it appear privileged because it is taken?"

Some sort of explanation in this sense for the principle of least action (it will be convenient continuing calling it thus) was finally provided by Feynman with his formulation of quantum mechanics in terms of *quantum path integrals* (Feynman and Hibbs, 1965).

Feynman's answer was basically of the following kind. The Newtonian trajectories are not part of fundamental physics—they are a statistical artefact. Consider the space of all possible quantum paths, each of which has a definite value of action determined by an action integral taken along the path. Let us navigate on an arbitrary course through the space of *paths*, all the while monitoring the action "dial," whose needle will of course go up and down as we move along. Associated with this dial there is a *quantum wave amplitude* meter that separately displays the magnitude and the phase of the current path's quantum amplitude (a complex number). The phase—so it turns out in quantum mechanics—is just the action wrapped around a circular scale, so that as the action goes up and down, the phase needle revolves around a circle, coming back to the same position after an entire turn. Whenever the action needle stops moving for a moment (for instance, when the action is about to reverse course) the phase needle itself will stop rotating—it will remain stationary—for a moment.

What is a neighbohood of a point in path space is better visualized as a "bundle" of paths in spacetime. Now, in quantum mechanics the probability of a bundle of paths is not proportional to *how many* paths the bundle contains—as one steeped in classical statistical mechanics might expect—but to the *absolute square* (in the sense of $|\psi^*\psi|$) of the sum of their amplitudes (Feynman and Hibbs, 1965). When, during our voyage in path space, the rate of change of action is high, in a brief space the phase will go through several turns. The quantum amplitudes of neighboring points will destructively interfere with one another and thus the probability of a bundle of neighboring paths will be close to zero.[3] Only when the rate of change of action vanishes—i.e., the action needle is *stationary*—will the amplitudes of neighboring paths add up constructively, yielding a nonvanishing probability for the bundle. Thus, spreading a Lagrangian blanket over $(q, \dot{q}, t)$ space is equivalent to spreading a probability blanket over path space. The latter blanket will generally lay low and flat, but will be crisscrossed by a web of high probability "ribs."[4] It is us who *single out* for special consideration these "probability wrinkles" on an otherwise flat blanket—and record them in our maps as Newtonian trajectories.

Thus, in Feynman's scheme of things, no "providential hand" is steering trajectories so as to minimize a certain "action expenditure." Rather, microscopic quantum paths uniformly cover all path space, but their effects cancel out everywhere except in certain distinguished regions. To "high-density"[5] neighborhoods in path space there correspond high-probability bundles of microscopic paths in spacetime. From a distance, these bundles may appear so narrow that we can idealize them as lines of zero width. It is these ideal lines that older physicists had been calling "Newtonian Trajectories."

There are two reasons why Feynman's "explanation" of the least-action principle may leave one somewehat less than satisfied.

The first reason is that the quantum path integral argument explains action—a quantity that was devised wholly within classical mechanics and had at least two centuries of independent currency in it—on the basis of quantum mechanics. Does that mean that, if quantum mechanics hadn't been arrived at yet, then

[3] More precisely, if N is the number of paths in the bundle, the probability of this bundle will be proportional to $\sqrt{N}$—rather than to N itself as in the classical case. The relative probability, proportional to $N/\sqrt{N}$, will vanish as N becomes large.

[4] An even better image may be soap-bubble foam, which is mostly low-density air except for a web of high-density soap-water walls, edges, and vertices (keep in mind, though, that path space has a fantastically large number of dimensions).

[5] In the absolute-square-sum-of-amplitudes sense seen above rather than in terms of mere count.

action would have to remain unexplained? Remark that there are in physics other principles of scope comparable to that of least action, and which also address phenomena apparently manifesting teleological or providential "forces," and which, however, unlike the least-action principle, are fully explained by internal logic, without involving postulates of an external nature. We have in mind, specifically, the *maximum entropy* principle (Jaynes, 1957), which fully accounts for classical statistical–mechanical phenomena purely in terms of classical mechanics, as well as quantum statistical phenomena in terms of quantum mechanics.

The second reason is that, besides being external—as it were—to classical mechanics, the quantum mechanics in terms of which the above path-integral argument claims to "explain" the least-action principle is itself *unexplained*—and it is Feynman himself that forcefully keeps reminding us of that (Feynman, 1963). Thus, we have reduced ourselves to explaining a *mystical* principle in terms of a *mysterious* principle. How much have we really gained? Richard Feynman himself, even after having stalked action—as we've seen—at closer distance than any other mortal, would still confess "I don't know what action is." Apparently, there are more veils to be lifted. By contrast, in spite of the veil of mystery that accompanied it through much of its life, today we can confidently say that we know what *entropy* is.

Be it clear that I do not claim to explain the least-action principle here. What I want to ask is, Did we really have to wait for quantum mechanics in order to have some explanation of this principle? Whether or not the explanation provided by quantum mechanics is the true and ultimate one, couldn't we formulate other plausible explanations within classical mechanics itself? This is what happened, for example, for thermodynamics, which appeared to introduce its own independent set of quantities and laws,[6] but ultimately was shown, via statistical mechanics—and thus by purely logical and combinatorial (that is, *not irreducibly physical*) arguments—to be but an epiphenomenon on top of ordinary mechanics (classical or quantum as one's level of investigation requires).

I will tell an edifying story here. In addressing the issue of the "spring of air" (what today we call Boyle's law for ideal gases, $pV = \text{const}$), Boyle (1660) compared air particles to coiled-up balls of wool or *springs* (what else?) which would resist compression and expand into any available space (Brush, 1976). Newton toyed with this idea and showed that first neighbors repelling one another *inversely as the distance* would give the required macroscopic behavior. What became popular next is models viewing air as a swarm of discrete particles freely jostling about and interacting with one another in some prescribed way, and Newton himself considered one with short range *attractive* forces. Bernouilli considered elastic collisions between hard spheres ("billiard ball" model), while Maxwell considered *repulsion* with an inverse power law—fifth-power being preferred for its computational advantages. Much more recently (1931), taking into account the

[6] cf. the Pure Thermodynamics of Mach and Ostwald (Brush, 1976).

quantum–mechanical electrical structure of orbitals, Lennard–Jones established, for simple nonpolar molecules an inverse seventh-power attractive law.

The moral of this story is that, if one is looking for a microscopic model for the laws of ideal gases, there is no dearth of *plausible* candidates. On the contrary, the same generic macroscopic behavior embodied by these laws will emerge from *almost any* microscopic model—using attractive or repulsive forces, power-law or exponential or whatever—that displays certain basic symmetries. In fact, even a model *with no interactions whatsoever* between particles—only collisions with the container's walls—will do perfectly well. The "spring of air" is the most generic macroscopic expression of certain conservation laws, not of the details of a particular microscopic dynamics. Conversely, the fact that many power-law repulsive models gave the right macroscopic results was no guarantee that the interparticle forces are repulsive, power-law, or, least of all, fifth-power. If there was an embarassment with the "spring of air" it was not that there was no plausibel model for it, but that there were too many to choose from.

Given the generality and *genericity* of the least-action principle, I'm amazed that the same predicament—having many plausible explanations and at worst not knowing how to pick the right one—did not occur. Instead, apparently no explanation at all was extant until the one based on quantum path integrals came on the scene. Virtually anywhere we turn—statistical mechanics, economics, operations research, population genetics, etc.—variational principles are obviously the surface expression of an *underlying fine-grained combinatorics*. On the other hand, virtually all books that I have seen that deal with the calculus of variations from a mathematical viewpoint (e.g., Denn, 1978; Forsyth, 1960; Sagan, 1969; Weinstock, 1974)—and these include the books on analytical mechanics (e.g., Aenold, 1978; Goldstein, 1980; Lanczos, 1970; Sussman *et al.*, 2001; Yourgrau and Mandelstan, 1979)—do not even *entertain the possibility* of a combinatorial origin. What's even graver, they do not mention that most variational principles happen to be of a *mathematical form* that, whatever its actual origin in each specific case, would most naturally *emerge* from an underlying fine-grained combinatorics.

In brief, I am reluctant to accept the law of least action as an unexplained first principle, or, for that matter, one whose *only* possible reduction were the quantum-mechanical "explanation". I'd be much happier if there were a range of plausible combinatorial hypotheses (one may imagine fine-grained, "sub-microscopic" mechanical models, or models where uncertainty about the dynamical laws or the state of the system's environment are taken into explicit consideration) from which the least-action principle would plainly emerge as a genuine statistical consequence—a feature of aggregate behavior. One would then have to decide which of these models satisfied additional requirements *besides* the least-action principle.

In this paper I will briefly consider a number of combinatorial approaches that present themselves as obvious "first-interview" candidates (whatever their ultimate merits) for the role of generator of the least-action principle. Whether the quantum–mechanical "explanation" should be counted among these depends on whether we think of quantum mechanics as a better (that is, more refined, more complete, and, at bottom, more precise) form of statistical mechanics (Youssef, 2001)—and thus a generic *inference theory* superposed on certain irreducibly physical primitives (cf. Jaynes' *Probability: The Logic of Science*; Jaynes, in press)—or rather itself part and parcel of those physical primitives, like most physicists are inclined to believe.

2. TYPICAL COMBINATORIAL ORIGIN OF VARIATIONAL PRINCIPLES

There is this summer camp with 100 kids. They are pennyless since they have spent all their money on their last visit to town. But on coming back to camp one of the kids finds an envelope with \$100 in one-dollar bills from his parents. For the next week they do not leave camp and no more money arrives. You know kids—they buy, sell, "traffic". By the end of the week the \$100 will be somehow redistributed among the 100 kids. What distribution do you expect?

You are not given any details about the dynamics of this system or its initial conditions. All you know is that money is *strictly conserved* and that 1 week is enough time for money to change hands innumerable times. In principle, *any partition* of this money, that is, any assignment $\langle n_1, n_2, \ldots n_{100}\rangle$ of n_i dollars to kid i, is possible, provided that

$$n_1 + n_2 + \cdots + n_{100} = 100.$$

One may maintain that there is no reason to prefer any one of these partitions to any other, in which case all the different partitions would be attributed the *same* probability. In turn, this probability can be justified in terms of a deeper argument: Consider the set of all possible combinations of *dynamics* and *initial conditions*, and for each derive the resulting partition of \$100 among the 100 kids. This an astronomically-many-to-one correspondence, the set of dynamics folding over the set of partitions innumerable times.

Thus, we are distributing dynamics over partitions, and the latter distribute dollars over kids. Unless *both* processes have perverse biases, the dollar expectation will be the same for all kids, namely, $\$100/100 = \$1$. The point we are going to make, trivial as it may be, is that this *expected partition*—a flat distribution of money over kids, as it happens—can be described indirectly by means of *variational principle*.

Let q_i be the dollar expectation for kid i. We demand a "trajectory" $q_1, q_2, \ldots, q_{100}$ such that

$$q_1 + q_2 + \cdots + q_{100} = 1 \tag{1}$$

and such that the integral over the trajectory of a certain *cost function* be stationary with respect to infinitesimal variations of the trajectory itself. The variations can be arbitrary, except that the variated trajectories must satisfy the normalization condition (1). As a cost function we will assign

$$L(q) = q^2. \tag{2}$$

Like the Lagrangian of a dynamical system, which need not be justified since it is what *defines* the system, so our cost function $L(q)$ need not be justified—it suffices that it generate the desired trajectory. However, our choice can be motivated on combinatorial grounds, the square function being essentially the $-$log of the binomial distribution. In other words, $-q_i^2$ expresses (up to an additive constant) the contribution given to the entropy of the distribution $\{q_1, \ldots, q_{100}\}$ by an an expectation q_i for kid i. In brief, we make a large value for q lead to a "costly" variation because it is *unlikely* in the underlying combinatorics. Note that this is no more and no less than the principle of least squares.

It turns out that the flat "trajectory", $q_1 = q_2 = \cdots = q_{100} \equiv 1$, is the one for which the overall cost $S = \sum_i L(q_i)$ is extremal. In fact, we require that

$$\delta S \equiv \sum_i L(q_i + \delta q_i) - \sum_i L(q_i) = 0 \tag{3}$$

subject to (1) and thus to $\sum_i \delta q_i = 0$. But, from (3),

$$\delta S = \sum_i 2q_i \delta q_i + (\delta q_i)^2,$$

which, if all the q_i have the same value $\overline{q} = 1$, reduces to

$$\delta S = \sum_i (\delta q_i)^2.$$

As a sum of squares, this is minimized when each individual term is zero.

If we turn now to a canonical example from the calculus of variations, that is, the issue of what shape is acquired under gravity by a chain stretched between two points, we find a combinatorial origin for the resulting shape (which happens to be a *catenary*) and a clarification of *what* precisely this is the shape of—matters about which variational textbooks are unaccountably silent.

Let a chain be suspended between points. At time $t = 0$ the chain is given a vigorous shake, so that it will start with the shape indicated in Fig. 1(a); each chain element will of course have a velocity besides a position, but that is not indicated in the figure. We assume ideal Newtonian mechanics—no friction, no air resistance, etc. Gravity is turned on, pulling downwards. We wait a long time and then we look again; what shape do we expect to see? Perhaps that of Fig. 1(b)?

Nonsense! The chain will have been quivering and snaking all the while, and, for all we know (which is little because the initial velocities were not recorded), it may now display any of the configurations compatible with the initial conditions; typically, we will see something like Fig. 1(c).

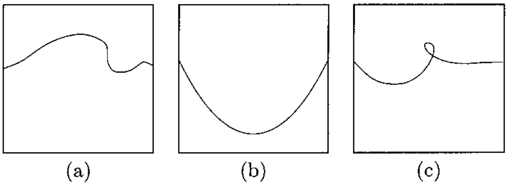

Fig. 1. (a) Initial configuration of a chain suspended between two points and given a vigorous shake. Though the chain is of course in motion at this moment, the velocity of the chain elements is not indicated in the figure. As is well known, the rest configuration of the chain is the catenary (b). However, in the stated regime (no friction) the chain will never come to rest configuration (c) is typical of what the chain may look like some time later. The catenary shape will be achieved only in the presence of damping, and then only *approximately* because, by its very nature, damping cannot be disassociated from some amount of wiggle.

In order to get something like a catenary we'll have to let air in and thus introduce a number of new degrees of freedom (a swarm of jostling particles) much larger than the number of chain elements. Under these conditions the energy of the chain will distribute itself evenly, on average, among all the available degrees of freedom, new and old, just as the $100 did in the summer camp.[7] The resulting chain configuration will not be strictly a catenary—since air and chain will keep interacting and forever disturbing one another[8]—but the *expected* configuration, that is, the mean of the microscopic configurations that make up the equilibrium ensemble, will be a smooth curve that can be made arbitrarily close to a catenary (Fig. 1(b)) by using a sufficiently large number of gas particles at a sufficiently low temperature.

Remark that

- No damping ⇒ No equilibrium configuration (never mind a catenary) will emerge.
- Damping ⇒ No rest can ever be achieved. The ever-changing microscopic configuration will not be a catenary, but the latter will be achieved *in the mean* (and then only in the limit for infinitely-fine-grained damping).

In conclusion, we see that the catenary curve prescribed by whatever variational principle is appropriate to the present case is not an actual configuration but a mathematical expectation obtained by taking the mean of a large family of

[7] Viewed from the standpoint of microscopic combinatorics, the theorem of *equipartition of energy* is but an admission of ignorance. In the summer camp case, we do not pretend to know that each kid will ultimately *end up* with $1; we only mean that that is each kid's expected share.

[8] See Feynman's admirable discussion (Feynman, 1963) on why damping is a Faustian bargain: the price of being driven towards equilibrium is never to be able to attain rest.

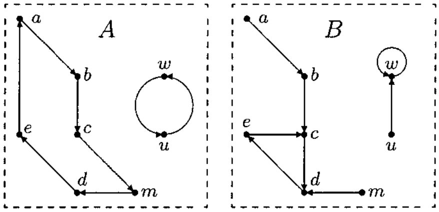

Fig. 2. Two discrete dynamical systems, each consisting of eight states. The first system, *A*, is invertible. The second, *B*, is not, since state *u*, for one, has no predecessors and state *c* has two.

microscopic curves. Whether the latter are actualized—for instance by letting the system run for a long time so that a time average can be collected—or are mentally contemplated as the set of all curves compatible with our lack of information, the catenary itself will be a statistical construct in either case. The situation is similar to that occurring in systems that rely on error for feedback, like a furnace/thermostat loop. The ideal "steady" temperature that is so achieved is actually the mean of a temperature that is continually ramping up and down.

Thus, when we read the fine print, we realize that the one "ideal" trajectory that the variational principles advertise turns out to be a blurred version of a large collection of "somewhat defective" ones.

In the summer camp example, the dissipative cascade that led to the emergence of a distinguished equilibrium distribution was initiated by the arrival of $100. In the catenary example, it was initiated by letting air—as a damping agent—into the system. How many ways are there to set up a dissipation/emergence engine?

3. WHAT COULD DRIVE AN INVISIBLE HAND?

3.1. Emergence by Microscopic Irreversibility

For us, a *dynamical system* is a just a map from a state set to itself.[9] The simplest way to see how preferred configurations or trajectories emerge in dissipative systems is to compare an invertible dynamical system with one that is not (see Fig. 2). By construction, each state has exactly one successor; if each state also has exactly one *predecessor*, then the system is invertible. In Fig. 2, for each state (denoted by a dot) the dynamics specifies (through an arrow) its successor. System *A* is clearly invertible. On the other hand, system *B* is noninvertible; for example, *u* has no predecessors and *c* has for predecessors both *b* and *e*.

Let the dynamics of systems *A* and *B* be publicly announced. Now I think of a specific initial state q_0 and ask you to guess it (hereafter, I will not give you any

[9] This for a discrete-time, time-independent system. For the sake of our argument, there is no need to consider more general cases.

feedback on your guesses). For either system, you have a 1-in-8 chance of guessing q_0 right; in other words, the entropy of the situation is for you 3 bits ($3 = \log_2 8$). We now "turn the crank," i.e., let the dynamics advance by one step, so that each state flows into its successor; specifically, q_0 will flow into a state q_1. I ask you to guess q_1. For system A, your uncertainty is still 1-in-8, but for system B it is only 1-in-5, because after the first step states a, m, and u are no longer possible, since no states flow into them. Similarly for the state q_2 reached after two steps, only four possibilities remain, namely c, d, e, and w, and in this example they all happen to have the same probability, 1/4. No further states are lost on subsequent steps, and in fact already by the second step the system has attained the invariant distribution $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ over those four states. Your initial uncertainty of 3 bits is thus ultimately reduced to 2 bits. The system has somewhat "ordered" itself, in the sense that you know more about its state now than you did at the beginning.

Where has this "new" information sprung from? Clearly, it was implicitly contained in the dynamics itself and it was "revealed" (like in a photographic developer bath) by running a certain number of steps. While invertible systems are by construction "information lossless," and thus preserve from step to step the amount of uncertainty about their current microstate, noninvertible systems typically "leak out" information, gradually reducing the amount of uncertainty about their state. Every time two trajectories merge, the system loses the information needed to know which of the two it came from; as we saw for system B, microstates that initially used to be accessible to the dynamics may from a certain point on become no longer accessible. If a noninvertible system started in a highly disordered macrostate eventually produces, say, a flower, it is because flower-like patterns, among all others, are preferentially retained by the dynamics as time goes by.

In Fig. 3, the middle panel shows a random initial state for a two-dimensional cellular automation. The left panel shows the state reached from there after 100 steps of an invertible dynamics σ, namely, an Ising spin model. The right panel shows the state reached after the same number of steps of a *noninvertible*

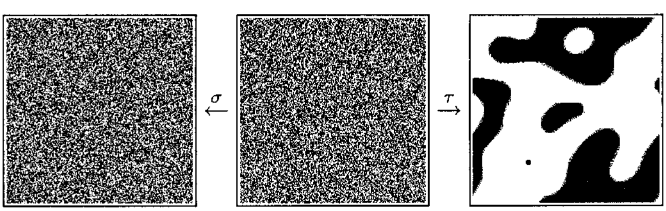

Fig. 3. We start from a maximally random initial state (middle). Running 100 steps of invertible dynamics σ—an Ising spin model—give us another maximally random state (left). On the other hand, 100 steps of *noninvertible* dynamics τ—a majority-voting rule—lead to the emergence of large homogeneous regions with characteristic boundary shapes.

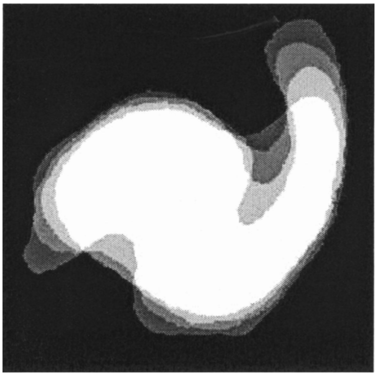

Fig. 4. With majority voting, small enclaves of one party get absorbed by the other party: the boundary between domains advances at a rate proportional to the *curvature*, thus trying to *minimize* boundary length. Here is a time-lapse photo combining four uniformly spaced snapshots.

dynamics τ, namely, a majority-voting rule; note the emergence of domains with boundary shapes characteristic of majority-voting's *annealing* behavior.

Because of the built-in tendency, with this voting scheme, for majorities to encroach on minorities, the boundary between adjacent domains moves with a speed approximately proportional to its curvature: smaller pockets are engulfed faster; this is illustrated by the time-lapse composite of Fig. 4. It is evident that boundaries move as if they were striving to *minimize* their overall length. We thus see macroscopic behavior that not only spontaneously proceeds toward greater order, but seems to do so under the guidance of a variational principle.

3.2. Emergence by Low-Entropy Initial Conditions

A related mode of emergence is that which occurs, *even when the underlying dynamics is invertible*, from low-entropy initial conditions. As soon as the dynamics' "crank" starts turning, this initial low-entropy state may start leaking into new regions of phase space, leading to a state of higher coarse-grained entropy.[10] The pattern or texture representative of this new state may look quite different from the original one. Specifically, microscopic correlations buried in the initial state may be converted into more visible macroscopic regularities.

This process of "directed evolution" will continue as long as the system can slide towards higher coarse-grained entropy, but will eventually slow down and grind to a halt as the system approaches equilibrium. Figure 5 shows a typical example of emergence of texture in an invertible system; we must remember that this is in reality a transition from hidden microscopic regularities to more evident coarse-grained ones.

[10] If the system is invertible, the *fine-grained* entropy will of course remain constant.

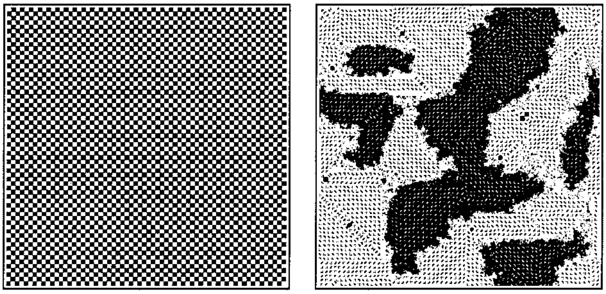

Fig. 5. Here the invertible microscopic dynamics is the same Ising spin model as used in the leftward path of Fig. 3. This time, however, we injected in the initial conditions some regularity at the microscopic level, in the form of a fine-grained checkerboard sprinkled with a minuscule amount of randomness (left). This regularity can be used as "fuel" to drive emergent behavior at a more macroscopic level (right). Note that, even though the microscopic dynamics remains that of the Ising model, the emergent dynamics is similar to that of the majority rule of Fig. 3 and, like the latter, tends to shrink domain boundaries as if driven by a variational principle. However, in this case the driving "engine" eventually runs out of statistical fuel and—unlike in the majority dynamics—the shrinking process does not go on indefinitely.

3.3. Emergence by Darwinian Evolution

In both of the previous sections, the emergence of characteristic patterns is obtained at the cost of *running down* irreplaceable "batteries"—namely, the "freshness" of the dynamics in the case of noninvertible systems, and the "orderliness" of the starting state in the case of a low-entropy initial configuration.

There may be circumstances, however, where a distinguished region of the system is effectively endowed with a permanent "power supply" instead of finite-life batteries—a mechanism whose ultimate effect is to drain off disorder even as it accumulates.[11]

An eminent example of power-assisted emergent dynamics is *life* on the surface of the Earth, which is sandwiched between the Sun—a steady source of predictable radiation—and the black cosmic background—an undiscriminating absorber of the reradiated thermal noise. Of course, this sharp dichotomy between light and darkness will not last forever; in the mean time, however, the thin, two-dimensional region where we live provides an ideal stage for the enacting of emergent dynamics.

[11] Typically, this is done by convection: a flow of high-grade energy is steadily pumped into that distinguished region, while low-grade, thermalized energy is pumped out.

The reader must be familiar with the fable of the investment consulting firm who sent a free advice newsletter saying, "Tomorrow IBM stock will go up" to 512 prospective customers, and "will go down" to another 512. Whichever way it turned out, the 512 prospects that had received wrong advice would be thenceforward ignored, while the "successful" 512 would be divided into two groups of 256 each. The mailing and subsequent culling would be repeated in a similar way for these two groups, and so on recursively eight times, at which point there would be only four prospects left—but these would have received an unbroken sequence of eight successful predictions. At this point the consulting firm would move in for the kill. Writing *only* to those four, they would say: "By now you must be convinced that we draw on very reliable proprietary forecasting methods. May we suggest that, for only $100,000 a year, you subscribe to our privileged weekly newsletter service?" (Not surprisingly, the next 52 issues of the newsletter will turn out to have only a 50% hit rate. What the heck—let the buyer beware!)

Regrettably, the above culling process had to shrink the pool of prospects from 1024 to 4, and clearly the remaining 1020 cannot be recycled for a new scam of the same kind. Wouldn't it be nice if we could make the latter group disappear from the face of the earth, and replace them with "clones" of the four that got all right predictions, thus amplifying their number from 4 to 1024? We could then, for example, stretch our preparatory mailing sequence by eight more accurate predictions, as if we had started from a base of 256k prospects rather than only 1k—but note that our costs will now be *linear* (rather than exponential) in the number of prediction stages.

This, in a nutshell, is what happens in *Darwinian evolution*. Its mechanism is a "power-assisted magnifier" that allows one to zoom in on what works best, blow it up to full scale, and sweep the rest out of the picture—and keep doing that over and over.

3.4. Emergence by External Conditioning of a Nanoscopic Dynamics

In the earlier three cases, the *dissipative culling* of configurations was done by means of resources *running within the physics*. In fact, in section 3.1 the one-time "batteries" came with the system itself, and the latter becomes useless as mechanism for driving emergence once these batteries have run down.[12] In section 3.2 the system came with no batteries; those had to be put in by the user in the form of a low-entropy initial state. When this source ran down, the emergence experiment would come to an end. In section 3.3 finally, the system came accompanied by a lifetime contract with a power company, so that one did not have to worry about running out of power.

[12] Just like disposable flashlights, which become useless for illumination purposes once their batteries are exhausted.

In analytical mechanics and in quantum mechanics, on the other hand, whatever culling is performed by the principle of least action cannot be driven by a dissipative cascade, since both theories purport to be conservative theories—what you see is what you get—not accounts of irreversible statistical epiphenomena. In other words, the world of analytical mechanics (as contrasted to *statistical* mechanics) never winds down (orbits can be traversed indifferently forward or backwards by "frictionless coastling"). Not only does it not need batteries to keep running but also it has no room for any.

In Fig. 1, the catenary was selected as a limiting mean shape out of innumerable alternative shapes by a physical process running in actual time: at one moment we see a shape that was definitely not a catenary and a little later we see a shape that is close to a catenary. If the least-action principle cannot make use of actual physical dissipation as a mechanism to select the desired mechanical trajectory out of innumerable alternatives, what other resources can be responsible for the selection process we wish to explain?

Our general plan is to have some underlying combinatorics do the dirty work on behalf of the least-action principle.

First of all, we'll have to summon the information–mechanical machinery within which this combinatorics will play itself out. It is tempting to postulate, for this purpose, an even finer-grained level of dynamics, which we may call *nanomechanics*, according to the following hierarchy

2	*macromechanics*	ordinary statistical mechanics
1	*micromechanics*	ordinary analytical mechanics
0	*nanomechanics*	the postulated finer-grained dynamics

Second, we must remember that the "culling" performed by the least-action principle operates on *trajectories*, not on *states*. Thus, while statistical mechanics establishes correspondence principles between collections, ensembles, or distributions of microstates from level 1 and constructs of level 2, generally called *macrostates*, we will be trying to interpret analytical mechanics as a theory of "aggregate behavior" establishing correspondence principles between collections or "bundles" of nanotrajectories from level 0 and the microtrajectories (the ordinary mechanical trajectories) of level 1. Note that Feynman's quantum paths belong to the same general class of constructs as these nanotrajectories.

Third, even though I greatly appreciate the value of the *maximum-entropy principle* (Jaynes, 1957) and the rationale behind it, for the present purposes I prefer to utilize a weaker, more tautological—and thus more compelling—version of it. Namely, if in a certain inference problem we are uncertain as to which macrostate out of a *range of possible macrostates* should be used for making statistical predictions, the maximum-entropy principle recommends that we use the one having the greatest entropy.[13] The weaker form of the principle I have in mind is that we should construct instead a new macrostate-like object from a *weighted average* of that range of macrostates (note that, unlike the maximum-entropy macrostate, this new object *need not* be itself a macrostate from that range), and use that object as a representative for the whole range; in this way, all of the elements of the range contribute *additively* to their representative. In any case, microtrajectories at level 1 are obtained as representatives of bundles of nanotrajectories from level 0. Note that Feynman's constructing an average amplitude for a bundle of neighboring paths and associating this average (and, at a later stage, its square) with the whole bundle—which is not itself a quantum path—belongs to the same general kind of approach.

Fourth, we have no qualms with the Bayesian approach of viewing probability as a "degree of belief" represented by a real number between 0 and 1 and continually updated as new conditioning information becomes available. However, we prefer to view the universe of potential outcomes (as well as the events, which are subsets of this universe) as *finite* sets—at least in principle. Instead of being introduced as a primitive concept, probability would then be a *derived* one, namely, the relative size (a ratio of integers) of an event.[14] This position eliminates the paradox of "what probability measure should we use to prime the Bayesian pump with?" since there would be only *one* probability measure to choose from.

Accordingly, our hypothetical nanomechanics—whatever its details—will have discrete configurations and discrete evolution rules, and each nanotrajectory that takes part in a ballot will have exactly one vote.

Finally, let us recall that the task of the least-action principle is to convert an implicit specification for a trajectory into an explicit one.[15] In Hamiltonian mechanics, the implicit specification is the initial *phase* $\langle q_0, p_0 \rangle$ and the explicit one is the corresponding phase $\langle q(t), p(t) \rangle$ at any later (or earlier) time t. The conversion is driven by a compressed but nonetheless rather transparent description of the dynamics, namely the *Hamiltonian* "lookup table" $H(q, p)$. To "uncompress" this description we use a "continuous algorithm" called Hamilton's equations: If we fix the "address' q and slightly move the other address from p to $p + \partial p$, the corresponding amount of change ∂H in the table's value explicitly specifies the rate of change of q, namely, $q = \partial H / \partial p$. Similarly (with just a change of sign) for the rate of change of p.

In Lagrangian mechanics, we implicitly specify a trajectory between two times t_0 and t_1 by giving *half of the information* in the form of the initial configuration q_0. For the other half of the information, instead of giving the velocity

[13] We are are skipping over some fine print here.

[14] Mine is by no means a frequentist position. Initial probabilities remain a *priori*, because they are part of the *definition* of a system (Cox, 1946).

[15] See Sussman, Wisdom, and Mayer (2001) for an admirable exposition.

(or momentum) at the same time $t = 0$ one gives the *final* configuration q_1. The conversion of this information to an explicit trajectory is driven by a compressed description of the dynamics, namely the *Lagrangian* "lookup table" $L(q, \dot{q})$. The "continuous algorithm" that "uncompresses" the Lagrangian by applying the least-action principle is called the Euler–Lagrange equation; this algorithm is a tad more complex than the one used for the Hamiltonian.

In either case, the given information ($\langle q_0, p_0\rangle$ or $\langle q_0, q_1\rangle$) completely specifies the intended microtrajectory out of a large set of potential ones. By introducing a *nanomechanics* one provides a much larger set of potential "nanotrajectories." After the "culling and weeding" of this set caused by supplying $\langle q_0, q_1\rangle$ (for example) as *conditioning information*, one may be left with a bundle of *many* nanotrajectories rather than just one. *All* of the nanotrajectories of this bundle would be associated with the sought-for microtrajectory.

An "explanation" of the least-action principle that we would be willing to accept as a real explanation is, for instance, one in which the numerical value *of the integral of the Lagrangian function over a microtrajectory—that is, the* action integral—*were shown to stand for the* logarithm[16] *of the* number of nanotrajectories *associated with that microtrajectory.*

4. A FEW QUESTIONS

Many papers in the past attempted to provide a material interpretation of the variational principles of mechanics, but of course tended to do so in terms of other continuum theories like wave-front optics. For example, in a paper written a few years after Feynman's seminal paper on quantum path integrals (Feynman, 1948), Landauer (1952) discusses certain aproximate correspondence rules between quantum mechanics and Hamilton–Jacobi theory and gives a pictorial discussion of the solutions of the Hamilton–Jacobi equation in terms of swarms of particles. However, these particles simply mark the direction and the velocity of the flow, like magnetic lines of force; their number is not meaningful.

In "Action, or the fungibility of computation" (Toffoli, 1998) we give a few toy examples of nanomechanics from which a definite micromechanics emerges along the lines sketched in section 3.4. Specifically, we consider the dynamics of a digital string that had been studied by the Information Mechanics group at MIT (cf. Margolus, 1988; Smith, 1994). From a distance, this looks like a classical elastic string, with position and velocity varying in a continuous way along the string. On a closer view, it becomes evident that the velocity is no longer a second independent parameter but, just like the overall shape, is an emergent aspect of the

[16] Of course up to a scale factor (including sign) and an additive constant; other gaging transformations may be appropriate.

fine-grained configuration.[17] For such a string to go through a certain motion it is necessary not only that the appropriate coarse-grained shapes succeed one another so as to produce the correct "movie" of the motion, but also that the fine-grained shapes, which determine the velocities, are compatible with the velocities that are observed. Thus, the "chances" (in terms of number of available nanotrajectories) to go from configuration P to Q and from there to R is not just the product of the number of nanotrajectories in the P-to-Q bundle times that in the Q-to-R bundle (as if we were estimating the chances of going from Boston to Washington in certain time by multiplying the frequency of the Boston-to-New York trains by that of New York-to-Washington trains); these "chances" also depends on how many paths of the first bundle are compatible, in terms of *fine-grained* shape on arriving at Q, with those of the second bundle, which depart from Q. By needing to have these "proper joining" constraints satisfied along the paths, a dynamics of this kind may achieve something akin to *phase* and *interference* by purely combinatorial means, without having to make recourse to complex amplitudes.

These example, however, are still too sketchy to provide a coherent theory of emergence of anaytical mechanics from a "nanoworld," or even a reliable sense of direction for such a theory.

A related tack, also mentioned in Toffoli (1998) is to observe that, in Lagrangian mechanics, each virtual path, even if not a trajectory of the given dynamics, is a solution of *some other* dynamics of the same general class. Counting virtual paths is thus a way to count *alternative dynamics*, or—which is the same—expressing the amount of uncertainty about *which nanodynamics is operating* that is left after observing, say, a small piece of legitimate microtrajectory.[18]

We thus can conceive of a hierarchy of "lacks of knowledge" in physics, each with its associated kind of statistical mechanics or, better, *inference theory*; namely

- Uncertainty as to *which state*. We all agree that that is what *entropy* measures.
- Uncertainty as to *which trajectory*. My guess is that that is what *action* really is. But trajectories are, after all, *computations*—ways to know what output q_1 will be produced by input q_0 once we assign a dynamics. In this sense, action measures *amount of computation* just as entropy measures amount of *information* (cf. Toffoli, 1998).
- Uncertainty as to *which law*. What physical quantity measures the amount of *this* uncertainty? Should we invent one?

[17] This may remind one of Mach's attempts to explain all potential energy as some form of "trapped" kinetic energy.

[18] Imagine observing two people playing chess. How sure can you be, after watching them for three moves, that they are actually playing chess rather than some other chess-like game? Given some metarules about what a game should be like, how many different games could they be playing?

Finally—and that of course is a Holy Grail that may stimulate exploration and adventure whether or not it actually exists—one may wonder whether, instead of quantum mechanics throwing light on the least-action principle, a combinatorial explanation of classical analytical mechanics may not help imagine a reductionistic interpretation of quantum mechanics itself.

5. CONCLUSIONS

In conclusion, the present kind of questions are important not so much as a way of justifying a posteriori, in the twenty-first century, the state-of-the-art structure of physics at the beginning of the nineteenth. Rather, the motivation is that principles of great generality must be by their very nature *trivial*, that is, expressions of broad tautological identities. If the principle of least action, which is so general, still looks somewhat mysterious, that means we still do not understand what it is really an expression of—what it is trying to tell us.

ACKNOWLEDGMENTS

This research was funded in part by the Department of Energy under grant no. DOE-4097-5.

REFERENCES

Arnold, V. (1978). *Mathematical Methods of Classical Mechanics*, Springer, Berlin.

Brush, G. (1976). *The Kind of Motion We Call Heat: A History of the Kinetic Theory of Gases in the 19th Century*, North-Holland, Amsterdam.

Cox, R. T. (1946). Probability, frequency, and reasonable expectation. *American Journal of Physics* **14**, 1–13.

Denn, M. (1978). *Optimization by Variational Methods*, Rober Krieger, Huntington, New York.

Feynman, R. (1948). Space–time approach to non-relativistic quantum mechanics, *Reviews of Modern Physics* **20**, 367–387.

Feynman, R. (1963). *Lectures on Physics*, Vol. I, Addison-Wesley, Reading, MA.

Feynman, R., and Hibbs, A. (1965). *Quantum Mechanics and Path Integrals*, McGraw-Hill, New York.

Forsyth, A. R. (1960). *Calculus of Variations*, Dover, New York.

Goldstein, H. (1980). *Classical Mechanics*, 2nd edn., Addison-Wesley, Reading, MA.

Hildebrandt, S. and Tromba, A. (1996). *The Parsimonious Universe: Shape and form in the natural world*. Springer, Berlin.

Jaynes, E. (1957). Information theory and statistical mechanics. *Physical Review* **106**, 620–630; Jaynes, E. (1957). Information theory and Statical Mechanics. *Physical Review* **108**, 171–190.

Jaynes, E. (2003). *Probability: The Logic of Science*, Cambridge University Press, Cambridge, UK.

Landauer, R. (1952). Path concepts in Hamilton–Jacobi theory, *American Journal of Physics* **20**, 363–367.

Lanczos, C. (1970). *The Variational Principles of Mechanics*, 4th edn., Dover, New York.

Margolus, N. (March 1988). *Physics and computation*, PhD Thesis, MIT/LCS/TR-415, MIT.

Sagan, H. (1969). *Introduction to the Calculus of Variations*, Dover, New York.

Smith, M. (May 1994). *Cellular automata methods in mathematical physics*, PhD Thesis, MIT/LCS/TR-615, MIT.

Sussman, G. Wisom, J., and Mayer, M. (2001). *Structure and Interpretation of Classical Mechanics*, MIT Press, Cambridge, MA.

Toffoli, T. (1998). Action, or the fungibility of computation. In *Feynman and Computation* A. Hey ed. Perseus, Reading, Massachusetts, pp. 348–392.

Weinstock, R. (1974). *Calculus of Variations—With Applications to Physics and Engineering*, Dover, New York.

Yourgrau, W. and Mandelstam, S. (1979). *Variational Principles in Dynamics and Quantum Theory*, Dover, New York.

Youssef, S. (October, 2001). Physics with exotic probability theory. arXiv.org, hep-th/0110253.