AIP

PHYSICS TODAY

Whence the Force of $F = ma$? II: Rationalizations

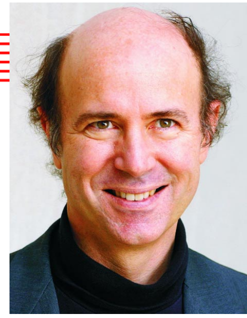
Frank Wilczek

Check for updates

Physics Today **57** (12), 10–11 (2004);
https://doi.org/10.1063/1.1878312

View Online
Export Citation
CrossMark

13 October 2025 07:34:01



Whence the Force of $F = ma$? II: Rationalizations

Frank Wilczek

In my previous column (PHYSICS TODAY, October 2004, page 11), I discussed how assumptions about F and m give substance to the spirit of $F = ma$. I called this set of assumptions the culture of force. I mentioned that several elements of the culture, though often presented as "laws," appear rather strange from the perspective of modern physics. Here I discuss how, and under what circumstances, some of those assumptions emerge as consequences of modern fundamentals—or don't!

Critique of the zeroth law

Ironically, it is the most primitive element of the culture of force—the zeroth law, conservation of mass—that bears the subtlest relationship to modern fundamentals.

Is the conservation of mass as used in classical mechanics a consequence of the conservation of energy in special relativity? Superficially, the case might appear straightforward. In special relativity we learn that the mass of a body is its energy at rest divided by the speed of light squared ($m = E/c^2$); and for slowly moving bodies, it is approximately that. Since energy is a conserved quantity, this equation appears to supply an adequate candidate, E/c^2, to fill the role of mass in the culture of force.

That reasoning won't withstand scrutiny, however. The gap in its logic becomes evident when we consider how we routinely treat reactions or decays involving elementary particles.

To determine the possible motions, we must explicitly specify the mass of each particle coming in and of each particle going out. Mass is a property of isolated particles, whose masses are intrinsic properties—that is, all protons have one mass, all electrons have another, and so on. (For experts: "Mass" labels irreducible representations of the Poincaré group.) There is no separate principle of mass conservation. Rather, the energies and momenta of such particles are given in terms of their masses and velocities, by well-known formulas, and we constrain the motion by imposing conservation of energy and momentum. In general, it is simply not true that the sum of the masses of what goes in is the same as the sum of the masses of what goes out.

Of course when everything is slowly moving, then mass does reduce to approximately E/c^2. It might therefore appear as if the problem, that mass as such is not conserved, can be swept under the rug, for only inconspicuous (small and slowly moving) bulges betray it. The trouble is that as we develop mechanics, we want to focus on those bulges. That is, we want to use conservation of energy again, subtracting off the mass–energy exactly (or rather, in practice, ignoring it) and keeping only the kinetic part $E - mc^2 \cong 1/2\ mv^2$. But you can't squeeze two conservation laws (for mass and nonrelativistic energy) out of one (for relativistic energy) honestly. Ascribing conservation of mass to its approximate equality with E/c^2 begs an essential question: Why, in a wide variety of circumstances, is mass–energy accurately walled off, and not convertible into other forms of energy?

To illustrate the problem concretely and numerically, consider the reaction $^2H + {}^3H \rightarrow {}^4He + n$, which is central for attempts to achieve controlled fusion. The total mass of the deuterium plus tritium exceeds that of the alpha plus neutron by 17.6 MeV. Suppose that the deuterium and tritium are initially at rest. Then the alpha emerges at .04 c; the neutron at .17 c.

In the (D,T) reaction, mass is not accurately conserved, and (nonrelativistic) kinetic energy has been produced from scratch, even though no particle is moving at a speed very close to the speed of light. Relativistic energy is conserved, of course, but there is no useful way to divide it up into two pieces that are separately conserved. In thought experiments, by adjusting the masses, we could make this problem appear in situations where the motion is arbitrarily slow. Another way to keep the motion slow is to allow the liberated mass–energy to be shared among many bodies.

Recovering the zeroth law

Thus, by licensing the conversion of mass into energy, special relativity nullifies the zeroth law, in principle. Why is Nature so circumspect about exploiting this freedom? How did Antoine Lavoisier, in the historic experiments that helped launch modern chemistry, manage to reinforce a central principle (conservation of mass) that isn't really true?

Proper justification of the zeroth law requires appeal to specific, profound facts about matter.

To explain why most of the energy of ordinary matter is accurately locked up as mass, we must first appeal to some basic properties of nuclei, where almost all the mass resides. The crucial properties of nuclei are persistence and dynamical isolation. The persistence of individual nuclei is a consequence of baryon number and electric charge conservation, and the properties of nuclear forces, which result in a spectrum of quasi-stable isotopes. The physical separation of nuclei and their mutual electrostatic repulsion—Coulomb barriers—guarantee their approximate dynamical isolation. That approximate dynamical isolation is rendered completely effective by the substantial energy gaps between the ground state of a nucleus and its excited states. Since the internal energy of a nucleus cannot change by a little bit, then in response to small perturbations, it doesn't change at all.

Because the overwhelming bulk of the mass–energy of ordinary matter is concentrated in nuclei, the isolation and integrity of nuclei—their persistence and lack of effective internal structure—go most of the way toward justifying the zeroth law. But note that to get this far, we needed to appeal to quantum theory and special aspects of nuclear phenomenology! For it is quantum theory that makes the concept of energy gaps available, and it is only particular aspects of nuclear forces that insure substantial gaps above the ground state. If it were possible for nuclei to be very much

Frank Wilczek *is the Herman Feshbach Professor of Physics at the Massachusetts Institute of Technology in Cambridge.*

13 October 2025 07:34:01

© 2004 American Institute of Physics, S-0031-9228-0412-210-2

larger and less structured—like blobs of liquid or gas—the gaps would be small, and the mass–energy would not be locked up so completely.

Radioactivity is an exception to nuclear integrity, and more generally the assumption of dynamical isolation goes out the window in extreme conditions, such as we study in nuclear and particle physics. In those circumstances, conservation of mass simply fails. In the common decay $\pi^0 \rightarrow \gamma\gamma$, for example, a massive π^0 particle evolves into photons of zero mass.

The mass of an individual electron is a universal constant, as is its charge.Electrons do not support internal excitations, and the number of electrons is conserved (if we ignore weak interactions and pair creation). These facts are ultimately rooted in quantum field theory. Together, they guarantee the integrity of electron mass–energy.

In assembling ordinary matter from nuclei and electrons, electrostatics plays the dominant role. We learn in quantum theory that the active, outer-shell electrons move with velocities of order $\alpha c = e^2/4\pi\hbar \approx .007\, c$. This indicates that the energies in play in chemistry are of order $m_e(\alpha c)^2/m_e c^2 = \alpha^2 \approx 5 \times 10^{-5}$ times the electron mass–energy, which in turn is a small fraction of the nuclear mass–energy. So chemical reactions change the mass–energy only at the level of parts per billion, and Lavoisier rules!

Note that inner-shell electrons of heavy elements, with velocities of order $Z\alpha$, can be relativistic. But the inner core of a heavy atom—nucleus plus inner electron shells—ordinarily retains its integrity, because it is spatially isolated and has a large energy gap. So the mass–energy of the core is conserved, though it is *not* accurately equal to the sum of the mass–energy of its component electrons and nucleus.

Putting it all together, we justify Isaac Newton's zeroth law for ordinary matter by means of the integrity of nuclei, electrons, and heavy atom cores, together with the slowness of the motion of these building blocks. The principles of quantum theory, leading to large energy gaps, underlie the integrity; the smallness of α, the fine-structure constant, underlies the slow motion.

Newton defined mass as "quantity of matter," and assumed it to be conserved. The connotation of his phrase, which underlies his assumption, is that the building blocks of matter are rearranged, but neither created nor destroyed, in physical processes; and that the mass of a body is the sum of the masses of its building blocks. We've now seen, from the perspective of modern foundations, why ordinarily these assumptions form an excellent approximation, if we take the building blocks to be nuclei, heavy atom cores, and electrons.

It would be wrong to leave the story there, however. For with our next steps in analyzing matter, we depart from this familiar ground: first off a cliff, then into glorious flight. If we try to use more basic building blocks (protons and neutrons) instead of nuclei, then we discover that the masses don't add accurately. If we go further, to the level of quarks and gluons, we can largely derive the mass of nuclei from pure energy, as I've discussed in earlier columns.

Mass and gravity

On the face of it, this complex and approximate justification of the mass concept used in classical mechanics poses a paradox: How does this rickety construct manage to support stunningly precise and successful predictions in celestial mechanics? The answer is that it is bypassed. The forces of celestial mechanics are gravitational, and so proportional to mass, and m cancels from the two sides of $F = ma$. This cancellation in the equation for motion in response to gravity becomes a foundational principle in general relativity, where the path is identified as a geodesic in curved spacetime, with no mention of mass.

In contrast to a particle's *response* to gravity, the gravitational *influence* that the particle exerts is only approximately proportional to its mass; the rigorous version of Einstein's field equation relates spacetime curvature to energy–momentum density. As far as gravity is concerned, there is no separate measure of quantity of matter apart from energy; that the energy of ordinary matter is dominated by mass–energy is immaterial.

The third and fourth laws

The third and fourth laws are approximate versions of conservation of momentum and conservation of angular momentum, respectively. (Recall that the fourth law stated that all forces are two-body central forces.) In the modern foundations of physics these great conservation laws reflect the symmetry of physical laws under translation and rotation symmetry. Since these conservation laws are more accurate and profound than the assumptions about forces commonly used to "derive" them, those assumptions have truly become anachronisms. I believe that they should, with due honors, be retired.

Newton argued for his third law by observing that a system with unbalanced internal forces would begin to accelerate spontaneously, "which is never observed." But this argument really motivates the conservation of momentum directly. Similarly, one can "derive" conservation of angular momentum from the observation that bodies don't spin up spontaneously. Of course, as a matter of pedagogy, one would point out that action–reaction systems and two-body central forces provide especially simple ways to satisfy the conservation laws.

Tacit simplicities

Some tacit assumptions about the simplicity of F are so deeply embedded that we easily take them for granted. But they have profound roots.

In calculating the force, we take into account only nearby bodies. Why can we get away with that? Locality in quantum field theory, which deeply embodies basic requirements of special relativity and quantum mechanics, gives us expressions for energy and momentum at a point—and thereby for force—that depend only on the position of bodies near that point. Even so-called long-range electric and gravitational forces (actually $1/r^2$—still falling rapidly with distance) reflect the special properties of locally coupled gauge fields and their associated covariant derivatives.

Similarly, the absence of significant multibody forces is connected to the fact that sensible (renormalizable) quantum field theories can't support them.

In this column I've stressed, and maybe strained, the relationship between the culture of force and modern fundamentals. In the final column of this series, I'll discuss its importance both as a continuing, expanding endeavor and as a philosophical model. ■

Rights & Permissions

You may make single copies of articles or departments for private use or for research. Authorization does not extend to systematic or multiple reproduction, to copying for promotional purposes, to electronic storage or distribution (including on the Web), or to republication in any form. In all such cases, you must obtain specific, written permission from the American Institute of Physics.

Contact the
**AIP Rights and Permissions Office,
Suite 1NO1,
2 Huntington Quadrangle,
Melville, NY 11747-4502
Fax: 516-575-2450
Telephone: 516-576-2268
E-mail: rights@aip.org**

13 October 2025 07:34:01