Mass, Force, and Energy: A New Formulation

Stanley James Speck

Part 1

The concepts of Mass, Force, and Energy are notoriously obscure. What is Mass? What is Force? What is Energy? These questions are still alive now at the opening of the twenty first century.

The three concepts are fundamental. They form a sort of three-legged stool upon which all of classical and modern physics rests, yet strangely enough, the exact way in which they should finally be understood is still open to question.

Richard Feynman characterized energy as "a numerical quantity, which does not change when something happens. It is not a description of a mechanism, or anything concrete; it is just a strange fact that we can calculate some number, and when we finish watching nature go through her tricks and calculate the number again, it is the same."

But what exactly does that number represent? According to Feynman, "It is important to realize that in physics today, we have no knowledge of what energy is. We do not have a picture that energy comes in little blobs of a definite amount."

Energy is often thought of as something which 'makes things move.' But does it make things move, or is it simply a measure of the motion that things have?

Energy is said to exist in two general forms, kinetic and potential. Kinetic energy is sometimes called 'the energy of a moving body.' It is the product of one half the mass of the body times its velocity squared. Potential energy can be computed as the product of a gravitational constant times the mass of a body times the height the body will fall. This is the gravitational 'potential energy' of a body near the earth's surface. These simple formulas and others can be used to compute quantities of the two 'forms' of energy, but what is the essential difference between the two? And how is that seemingly odd notion of *potential* energy to be understood? Can there be energy which is 'potential' but not 'actual?' All of this is still obscure.

Energy is sometimes defined as the 'capacity to do Work.' James Clerk Maxwell described Work in this way: "Work, therefore, is a transference of energy from one system to another; the system which gives out energy is said to do work on the system which receives it, and the amount of energy given out by the first system is always exactly equal to that received by the second." Incorporated in this description is the well known 'conservation of energy' principle. Yet Work is here described in terms of energy transfer. If our goal is to understand the nature of energy, it does us little good to characterize Work in terms of energy and energy in terms of Work, a circular conundrum. Sir Joseph Larmor, mindful of that dilemma, added this clarifying footnote: "The work done is a quantitative measure of the effort expended in deranging the system in terms of the consumption of energy that is required to give effect to it." If this comment is at all helpful in understanding 'Work,' it is due that essential word 'effort.' Work is defined as a 'quantitative measure of effort.' Yet the concept of 'effort' is nowhere defined as a physical concept. Larmor seems to have tacitly introduced an apparently anthropomorphic concept. All of this is evidence of how difficult it is to construct a non-circular definition and to avoid importing anthropomorphic notions of agency. Physicists often speak of bodies 'acting upon' other bodies. The notion of 'agency' seems to be critical to our usual understanding of the concept of 'force' (and by extension 'energy'), yet the concept of 'agency' (if it is a physical concept) is left vague and undefined.

Surely the most famous energy equation is $E = mc^2$, the epic-making linking of mass and energy. What seems especially peculiar and unexpected is not simply the apparent reduction of mass to energy but the incorporation in that equivalence of a definite finite speed, the speed of light. The suggestion that energy ultimately has something to do with a specific speed (c) and a specific constant (c^2) seems particularly puzzling.

It could be argued that all physical concepts, and all concepts of any kind, are ultimately obscure at some deepest level, that in some Hegelian sense every concept must be defined in terms of other concepts and those concepts in terms of still others. We can never become completely clear about the meaning of any single concept, so this argument goes, without engaging some larger, ever expanding "totality" of concepts. Whatever the merits of this argument (and I think it has some) we need not pursue it. Our aim is more limited and concrete: can energy be characterized in some way which is both useful and illuminating? Can it be made more accessible to the understanding than is currently the case?

Certain fundamental physical concepts seem intuitively clear and arouse little if any controversy as to their nature. These include, for example, the concepts of 'speed,' 'velocity' and 'acceleration.' Physicists are not likely to ask 'what is speed?' or 'what is velocity?' in the way they might ask 'what is mass?' Speed is the rate at which something is moving relative to some frame of reference. Velocity is the speed at which something is moving *in a certain direction* relative to some frame of reference. Acceleration is the rate at which a speed or a velocity is changing, namely, how fast something is speeding up or slowing down. These concepts are familiar and seem unproblematic. Of course, our intuitive grasp may be misleading. We know that these concepts depend upon the more fundamental concepts of distance and time, and no one since 1905 can regard even 'distance' and 'time' as unproblematic. Still, our experience with speeding automobiles and flying baseballs gives us an intuitive sense of what it means to say that something has a certain speed or that it is undergoing acceleration.

Is there some way in which the concepts of mass, force, and energy can be made similarly accessible to the understanding? Can they be given a new formulation which strips away some of their mystery? Is there a concept, perhaps surprisingly simple, which will explain and unite them?

The answer offered here may seem at first sight implausible or perhaps even absurd, but our effort is speculative and therefore requires a certain leniency on your part. Speculation is the father of invention. It is not required of speculation that it provide a fully satisfactory or well developed theory. It seeks, at the minimum, a pathway not a destination. Yet it carries with it the responsibility to be definite and not hopelessly vague. To paraphrase Wolfgang Pauli, the ideas must be clear enough and definite enough to be wrong. They must also be rich enough to provide a clear path to development. Such is our aim.

Consequently, we begin this endeavor by noting that the modern science of physics beginning with Galileo and his contemporaries seems to be primarily a study of motion. The careful observation of the motion of bodies rolling down inclined planes allowed Galileo to abstract precise relationships between distance and time that laid the foundation for classical physics. Newton extended these ideas and formulated his remarkable 'laws of motion' which became the basis for all subsequent inquiry. The development of thermal mechanics in the nineteenth century brought the concept of 'thermal energy' into the picture and ultimately led to energy's preeminent role in the science. Statics, as one physicist put it, is merely a special case of dynamics. One might say with little exaggeration that physics is 'the science of motion.'

James Clerk Maxwell wrote a little book called 'Matter and Motion' in which he succinctly sums up the core notions as he saw them. He was a great synthesizer and had an uncanny ability, like all great thinkers, to abstract essential notions from the inessential chaos of experience. He did not go so far as to say that 'matter *is* motion' – surely a paradoxical or even a seemingly far-fetched idea – but he did focus on and understand the primacy of motion.

Yet in spite of all this emphasis on motion, there is one concept which might be said to be conspicuously absent. It is the elephant not in the room. We use concepts like speed, velocity, and acceleration to measure the movement of things, but they measure only *rates* of movement and not what we might call 'quantities of movement.' If motion is the subject of our study, and if our study is a rigorously quantitative one, the single concept which might be expected lie at the center of this endeavor is the concept of 'quantity of motion.'

The term 'quantity of motion' cropped up early in the modern history of physics when Descartes used it to refer to what we now call kinetic energy. 'Quantity of motion' has also been used to label or describe Momentum. Both descriptions are essentially correct. But the insight contained in this choice of terminology was never developed. Perhaps for good reason, for it makes sense to say that objects have motion, or that matter has motion, or even that "masses" have motion, but to say that matter *is* motion seems nonsensical.

Yet this strange and startling idea is our subject. We hypothesize the ontological primacy of motion. In our ontology, motion is not an 'accidental' property of matter but the essential constituent. We assert the ultimate reduction of matter to motion. "Quantities of motion" are thus not missing in the centuries-long enterprise of natural philosophy; they are there but in disguise. The concepts of Mass, Force, Energy, Momentum, and Work all relate to the quantification of motion. Energy is a "quantity of motion."

What is a quantity of motion? It is, to be somewhat more precise, *the sum of the speeds of all the 'particles' in the system*. This idea will require considerable refinement, but at this juncture, let us change our approach and proceed by way of illustration. This will be the most effective and dramatic way to convey the essential ideas we wish to put forward. We will begin with a rather crude but simple illustration in order to convey the ideas in as simple and direct a manner as possible. We will then incrementally refine these ideas by adding the complexities which would be difficult to handle all at once.

Consequently, you are asked to imagine a billiard table and a set of billiard balls. We stipulate that the billiard balls are the 'same' in every respect relevant for our purposes (weight, size, and shape). We further stipulate that the table is frictionless and the balls and table are perfectly elastic. We do this in order to eliminate complicating and irrelevant variables. There are no such things as frictionless tables and perfectly elastic billiard balls (except in the domain of the imagination) but that should not invalidate our "thought experiment." If there were such things, we can inquire as to how these moving bodies might behave. Having accomplished that, we will extend our analysis to account for friction and elasticity. But initially, we simply want to stipulate that no energy is to be gained or lost once the balls are set in motion.

Under these somewhat artificial conditions, let us say that a billiard ball is struck with the cue stick and accelerated to a speed of ten meters per second. It will cross the table, rebound, and return with the same speed that was initially imparted to it. Since no energy is gained or lost, the ball will continue in motion indefinitely, crossing the table with the same exact speed: ten meters per second. Since no energy is gained or lost, the speed of the ball is "conserved."

Now let us add one further condition to our experiment. We ask you to imagine that this table exists in an 'alternative universe' where the laws of physics are somewhat simpler than in our own. You might object that this is a rather large concession; anything seems permissible in such a so-called universe. But our purpose is merely instructional. Once the central concepts we wish to illustrate have become clear, we will adjust the conditions back to our own universe to discover in what way they can be adapted to a real world.

So this time we place several balls on our frictionless table in our fanciful world. We strike the cue ball and again impart to it a speed of ten meters per second. The ball travels a short distance and strikes a second ball slightly off center. Immediately after this collision we measure the instantaneous speeds of the two balls now in motion, the cue ball and the struck ball. We find that one is traveling six meters per second and the other four meters per second. Each of these two balls then travels a short distance and strikes another ball. Again we measure the speeds of the four balls now in motion. The speeds are four, two, three, and one meters per second. A few moments later we again measure the speeds of all the balls in motion. We find that eight are in motion: six are traveling at one meter per second and two at two meters per second. (The nice round numbers used in this illustration are merely for ease of computation and have no other significance.)

Now we discover in our fanciful universe an interesting phenomenon: the *value* of the speed of the initial ball is *conserved* (given the conditions of our experiment). Whenever the speeds of the balls in motion are *summed*, we arrive at the same number, in this case, ten. Yet, except for the initial instance of the cue ball, this number does not represent the speed of any single ball. Subsequent to the cue off, it does not represent a speed at all. It is a combination of speeds. What it represents, we suggest, is a 'quantity of motion.' In our fanciful universe, we call it the *energy* of the system.

So we say that the energy of the system is the sum of the motion of all the particles in the system (the particles in this case are the billiard balls), and we draw this conservation law: the sum of speeds of the (ultimate) particles in any isolated system is always conserved. This is the conservation is energy.

Energy is not, consequently, some invisible and mysterious "I-know-not-what." It is not some vague essence or hidden "something" which causes things to move. Energy is motion itself. Or more properly stated: it is a quantity of motion. Energy is visible in the motion of things, but only by careful measurement and abstraction can its quantification be understood.

We noticed that the quantity of motion represented by the sum of the speeds of the balls on our hypothetical billiard table was conserved, but also conserved in a particular way. That is to say, the speed of the initial ball became 'spread out' or shared amongst the other balls. We began with one ball traveling at ten meters per second and ended with eight balls with speeds in the low digits. No matter how long we observe this system, it would be highly unlikely that we will again see a single ball moving at ten meters per second due to probability. The initial motion is 'shared out.' The 'sharing out' is a concomitant of the fact that when balls collide, they will interact, that is, motion will be divided to one extent or another. In time, all or most of the stationary balls will be struck and will begin to move. In time, they will all move at more or less the same speed (if never exactly the same speed). The likelihood for the initial motion to dissipate amongst the balls due to probability of collision is something we label 'entropy.'

Now let us return to the billiard table and confront this situation: a billiard ball traveling at two meters per second due east (as we shall label this direction) collides head-on with a billiard ball traveling in the opposite direction, due west, at the same speed. The two balls (equal in weight) rebound equally, reversing direction due to the fact that they are perfectly elastic. They travel away from each other at the same speed that they approached. The motion is ultimately conserved. It is as if the motion had traveled on through the two balls or as if the two balls had "passed through" each other, metaphorically speaking. The motion was perfectly exchanged.

Nevertheless, there seems to be a significant problem here. For we know that the two balls, in order to exchange their motion, had to come to a complete stop, if only for an instant. If they came to a complete stop to reverse the direction of their motion, what happened to their motion in that brief moment?

We have hypothesized that motion is strictly conserved. We mean by this that motion, the total quantity of motion in any (hypothetical) "isolated" system, is a fundamental and irreducible entity. As a sum, it never shows gain and never shows loss. Motion is exchanged but not augmented or diminished. It is not an accidental quality, but primary and essential, a universal constant. But if this is the case, how can we account for the missing motion at the brief moment of collision? Where does it go?

The answer is that it is converted to 'Mass.'

When the two balls initially slow down on impact, their former translational motion is converted to mass which increases. The mass of the balls increases until the balls come to a complete stop. Then, as the balls rebound and accelerate, the mass decreases correspondingly. The mass 'gives up' the motion which is returned to the balls in the form of translational motion across the table, or as physicists would say, in the form of 'kinetic energy.'

What, then, is this thing called Mass? What is this thing which receives and gives up motion?

The answer can only be one thing: mass is itself a 'quantity of motion.' It is the same thing as energy. It is, in other words, a sum of the motion of particles. There is, however, this one difference: mass is a quantity of motion *bound or confined to a certain 'region of space.'* The difference is primarily a matter of how it is measured.

The relevant 'region of space' in this particular case is the billiard ball itself. We must think of the billiard ball in our fanciful world not as a solid block of 'substance' but as a region of space which contains a vast quantity of motion within it. It contains an extraordinary multitude of particles and particles-within-particles, all in motion. It is a vast bee hive of motion. And the sum of this motion contained in the ball is what we call its mass. This motion includes that of the molecules as they agitate (thermal motion), the motion of the electrons, the motion of the constituents of the electrons, and most significantly, the vast motion contained in the interior of the nucleons.

Consequently, when the ball rolls across the table, each particle within the ball is given a very slight incremental increase in its translational motion. Compared to the motion contained in the ball itself, this incremental increase contributed by the ball's translational movement across the table is infinitesimally small. Consequently, the increase in the mass of the balls in collision is similarly small at normal billiard ball speeds. It would take a very delicate process to measure the increase in mass of the two balls at their moment of collision as compared to two stationary balls not in collision. The motion delivered to the system by the cue can thus be measured as mass or computed as kinetic energy. In both cases, what is being measured is same thing:

energy - which is to say, the total quantity of motion in the system.

If we think of billiard balls (and all other massive objects) as containing quantities of motion, then it is apparent that the moving billiard ball is a ‘quantity-of-motion in motion.’

Now let us backtrack for a moment. Consider the case of the collision of a billiard ball traveling at two meters a second (on our fanciful table) with another ball stationary on the table. Suppose the entire translational motion of the moving ball is transferred to the stationary ball. Since the two balls are identical, and since the moving ball loses its motion completely, and since the stationary balls gains the entirety of this motion, it seems intuitively persuasive to believe that the second ball will then accelerate to a speed of two meters per second. The quality of motion in this case is conserved. The speed (two meters per second) is simply transferred from the first ball to the second.

Now let us work a slight variation on this case. Suppose instead of a single stationary ball, two balls are stationary but linked together by some means. Now when this two-ball combination is struck by the single ball traveling at two meters per second, what will be the final speed of the two-ball combination? The answer seems intuitively clear and follows implicitly from our initial formulations. The final speed of the two-ball combination will be one meter per second.

When a billiard ball with a mass of one unit traveling at two meters per second strikes a second, stationary billiard-ball-combination with a mass of two units, the second ball (combination) will move at one meter per second if all the motion of the first ball is transferred to it. (This is simply equivalent to two balls traveling at one meter per second.)

Now let us imagine that the two balls are compressed into the space of one ball. A billiard ball made of heavy wood, twice as dense as a billiard ball made of light wood, will act ‘as if’ it were two balls as far as our computations go. The heavy ball has twice as much motion contained within it, and will thus move only half as fast when the same amount of motion is transferred to it as opposed to the lighter ball. Thus we multiply ‘mass times speed’ to find the correct quantity of motion.

We can compute mass in “billiard ball units” if we choose. We can also label these units as “kilograms” or “pounds”. The label is not important but we must be careful to keep the proportions consistent. For instance, let us specify that each of our ‘billiard-ball-units’ corresponds to one tenth ‘kilogram’ units, and one kilogram unit corresponds to 2.2 ‘pound’ units.

Initially, we stipulated that all the balls in our experiment were the same weight and thus the same mass. But it is now apparent that for balls of different mass, we need only compute the ‘units of mass’ in each ball to compute the ultimate quantity of motion, whether we choose to label these units “kilograms,” “pounds,” or “billiard-ball-units.”

We hypothesized that when the two billiard balls collided head on and came to a complete stop, the translational motion of the balls (their ‘kinetic energy’) was conveyed to the interior of the balls and thus became added mass. Yet it was not “permanent” mass. It was immediately “given up” in the form of kinetic energy when the balls rebounded. But at the moment before the motion was given up, it can be termed “potential energy.” Potential energy, loosely speaking, is potential motion but not in the sense that the motion comes from something other than some other motion in the system. It is simply a transfer or transmutation of one form of motion into another. The motion contained in the two balls (momentarily stationary under impact) is transferred or transmuted into translational motion when the balls rebound and accelerate. The difference between potential energies and other energies is again mostly a matter of how we go about our measurements. "Potential energy" can be regarded as 'hidden motion' in that it is often not obvious where the motion resides when it is hidden. But hidden does not mean nonexistent. Consequently, the distinction between potential and "actual" energy is somewhat artificial although this does not diminish the utility of the concept of 'potential energy' in our computations. (Interestingly, the idea that the motion is not lost, but simply internalized in collision, goes back at least to Leibnitz.)

It is obvious that the transfer and transmutation of motion in this scheme (or in any other) is a complicated business. Why do the two billiard ball absorb and then quickly 'give up' most of their motion in the collision? We know much about these interactions, but the notion of 'elasticity' is still not fully understood. It is usually understood in terms of 'forces' acting upon the bodies in motion. However, in our formulation we will draw a sharp distinction between 'forces' as agents and Force (denoted here with the capital 'F'). For us, Force refers a calculable quantity. Given our formulation, energy does not cause things to move any more than 'velocity' causes things to move. And given our formulations, 'Force' does not cause things to move any more than 'acceleration' causes things to move. Nevertheless, when motion is exchanged, we can measure the accelerations and the Force. Thus it is apparent that Force must be understood as the 'rate of change of a quantity of motion.' In the language of calculus, it is the derivative of that quantity of motion which we described as 'mass times speed.' Consequently, Force is mass times acceleration.

To understand what we mean here, consider the acceleration of the cue ball. When the ball is accelerated, a certain quantity of motion (measurable as mass times speed) is transferred to it from the cue. However, the rate of transfer can vary. If the cue ball is accelerated slowly to its final speed (say, to ten meters per second) by a lazy player, the rate of transfer is less than if the ball is accelerated quickly, even if the same final speed is obtained. It is apparent that Force does not "cause" things to move in this interpretation. Rather, it is simply used to describe the *rate* of exchange of motion. As such, it is a useful and essential concept. However, we have capitalized the term 'Force' to distinguish this concept from 'forces' - those hypothetical entities which are said to cause attractions and repulsions, that is, to cause exchanges of motion. We will look into this idea later. For our present purposes, we need only note that for whatever reasons motions transfer, the rate of transfer can be denoted with the concept of Force, equivalent to mass times the acceleration of the body.

Nevertheless, there are certain general and evident conclusions (given our formulations) that can be made concerning the question of what causes things to move. Given our formulation, things move when motion is added to them. And things stop when motion is taken away. We are interested here in finding the 'rules' or 'laws' (the 'logic') by which these transfers of motion take place. An object in motion will continue in motion until the motion is taken away and transferred to something else because if it stopped without such a transfer, motion would have to disappear from the system (go out of existence) which is impossible. Similarly, an object cannot begin to move unless motion is added to it from somewhere else, for if it did, the motion would have to come into being from nothing, which we deem impossible. If the quantity of motion in any isolated system is an inalterable sum, which is our interpretation of the conservation of mass and energy, then to suggest that something can begin to move without motion from somewhere else being transferred to it, would be equivalent to an effect without a cause. Consequently, things do not "resist" moving, nor do they "resist" stopping. A locomotive does not resist moving any more than a billiard ball. When we say that the locomotive has more 'inertia' we simply mean that the equivalent amount of motion added to the locomotive will impart a smaller speed to it than to a single billiard ball because the locomotive has more units of motion which must be accelerated. (This is equivalent to saying that a certain amount of motion transferred to a single billiard ball will result in a greater subsequent speed than if the same amount were divided among thousands of balls. The locomotive contains the equivalent of thousands of billiard-ball-units of motion.)

Inertia is sometimes interpreted as a 'something' which 'causes' a body to stay at rest and which must be 'overcome' in order for the body to move. Curiously, inertia is also sometimes thought of as something which keeps a body moving once it is set in motion. There is no such 'something' in our formulation, unless one simply means by it, that motion must be supplied to, or removed from a body for it to move or to stop moving.

Now let us eliminate some of the restrictions placed on our experiment in order to complete this preliminary sketch. If the billiard table is not frictionless, some motion will escape from the ball as it moves across the table. A certain increment of motion will be transferred from the moving ball to the molecules in the table top (and to the molecules in the air above it). Similarly, if the balls are not perfectly elastic, some of the motion of collision will be retained in the balls as increased movement of the molecules and consequently as heat. We can measure the amount of heat in an object with a thermometer. The thermometer reading indicates a certain level of motion in the molecules of the substance. When a quantity of motion is collectively removed from the substance, that is, when a given amount of motion is taken away from the molecules and transferred somewhere else, the thermometer will register a lower temperature. Temperature is thus a proxy for the motion of the particles in the object. The differences between the two temperature readings is a proxy indicating the how much motion has been removed and applied to some other system, in other words, how much "Work" has been done (and how much 'energy' (motion) has been transferred).

There are other potential losses or gains in motion. A billiard ball continually radiates motion away from itself in the form of photons which travel out from its surface. It also continually absorbs similar amounts. If the absorption exceeds the radiation (say, when placed under a heat lamp), the ball heats up, which is to say, it gains internal motion. Similarly, if radiation exceeds absorption, the motion lessens and the ball cools.

If mass is a quantity of motion (i.e. a quantity of energy), then it has a relative nature because motion itself is relative. If you accelerate a billiard ball relative to some framework (say, the billiard table), then the ball will acquire mass by virtue of that motion in proportion to the rest mass of the ball. The 'rest mass' does not change. This is to say that the relative motion of the particles within the ball does not change. If you were moving along with the ball at the same speed, its mass would not change, relative to you, because you would observe no additional motion. However, the quantity of motion represented by the motion of the ball across the table is equivalent to additional mass *measured from the framework of the table*.

At ordinary speeds the translational motion of the ball across the table makes very little difference to its mass (measured from the framework of the table). But if the ball were to be accelerated to very high rates of speed, such that the translational motion contributed a sum of motion comparable to that of the motion of the particles within the ball itself, a significant mass increase would occur relative to the stationary observer. We will return to this complicated subject later, but we need postulate no fundamental difference between 'rest mass' and 'relative' or 'kinetic' mass in this formulation. Rest mass is a consequence of motion within the object (bound motion) and therefore 'relative' in the sense that all motion is relative. The ultimate difference between 'rest mass' and 'relative mass' is a difference which primarily concerns the procedures of measurement.

This brief and preliminary look at our new formulations cannot be complete unless we touch upon a question which will doubtless have occurred to you. To say that energy (or mass) is a quantity of motion does not obviate the need for a 'something' which moves. What could that ultimate something possibly be? What is this entity which moves?

We must postulate that it is the smallest conceivable object. We call it a 'particle-point.' The 'particle-point' is similar to the 'geometric point' but unlike the geometric point, the particle-point exists in time. Its essence is motion. It is unobservable and has no inherent (or 'intrinsic') properties. It is dimensionless, and like the geometric point, knowable only by inference. But unlike the geometric point, it does not exist in any precise or specific 'place,' for it exists only in motion. And while there are an infinite number of geometric points in any region of space, there are only a finite number of particle-points representing the energy density.

Our task is to seek out the principles or 'rules' which govern this world, to find the 'axioms' or the 'logic' of space and time, much like the mathematicians who discover the axioms of a geometry. In a geometry there are geometric points, lines, and surfaces and volumes. In physics there are particle-points, paths, and material objects. The analogy is close, but not too close.

The postulated 'particle-points' are found on the sub-subatomic level, if you will pardon that expression. This notion of particle-points would have seemed most likely preposterous prior to the twentieth century. Before then, most physicists took it more or less for granted that nature 'at the bottom' must consist of atoms which were conceived of as tiny units of "matter." For Newton, atoms were "small, massy, and impenetrable." The question of what, precisely, these impenetrable units are made of haunts this formulation. In the twentieth century, the notion of an impenetrable 'substance' more or less gave way to a flux of mathematical equations even if it was never precisely clear what these equations ultimately represented. Yet, underlying even these equations is the smallest of all conceivable units, the geometric point.

There is more to be said about this subject, but before we venture there, let us return to our billiard table and bring it into the 'real world' as promised. While many of the basic ideas which we postulated to hold true for our fanciful world will be retained, others must be revised.

Let us, then, relocate the billiard table to the real world and run our experiment again. In doing so, we must introduce an essential element of the real world which was neglected or glossed over in our fanciful world: the element of 'direction.'

In our fanciful world, the number '10' represented in our scheme the collective motion of the moving balls, the sum of their speeds, a quantity which was conserved. We know that in our real world, this number is also conserved but not in exactly the same way. In our fanciful world we neglected to take into adequate consideration *how* bodies move. Bodies do not simply move; they move in a certain direction. They move either this way or that way. And the way – the direction – in which they move is critical. In order to sum up a quantity of motion, even to describe it, careful note must be taken of direction.

From any given point on the billiard table (or anywhere else) an infinite number of directions can radiate. A billiard ball at the center of the table can potentially move in any direction if given the appropriate impetus. A billiard ball is normally restricted to two-dimensional motion if it is not to fly off of the table, but we can easily imagine motion in all other directions as well. Suppose you are standing in the center of a large field equipped with a handgun and unlimited ammunition. You can fire off bullets in an unlimited number of directions (assuming you have unlimited time), including downwards towards your feet and upwards into the sky. Each bullet can travel in a unique direction from the center of the field outwards. Each bullet represents a small mass given a speed by the explosion in the gun. The combination of the two, the quantity of mass in the bullet multiplied by the value of the speed, represents in our scheme a quantity of motion *in that particular direction*. It is also called Momentum.

We have no difficulty in imaging these bits of mass being flung out into space from a certain position in the field. However, it difficult to imagine, and seems deeply counter-intuitive to suppose that motion has not been created and added to these bullets. It seems difficult to imagine that the motion was always 'there' and always there *in that particular direction*. Surely, the powder which exploded in the bullet must have created this motion which now expresses itself as an object flying rapidly skyward? This is certainly a reasonable interpretation, but not true as it happens. Motion in that particular direction was not created, but only 'borrowed' once we take the recoil of the gun into consideration. We can say that it was simply 'reassembled' or transmuted into a different configuration. Odd as this may sound, the quantity of motion in every direction is strictly conserved, one of the most arresting insights of classical physics.

To make sense of this claim, let us first ask: 'how would it ever be possible to keep track of quantities of motion in an infinite number of directions'? This daunting task would surely defeat the efforts of any man or machine. But Nature has made the task immensely easier. She has allowed us to configure all directions in terms of only six. We will label these directions: east, west, north, south, up and down. These, in terms, can be resolved into three familiar axis's: east-west, north-south, and up-down, the three "dimensions" of our real world. It should be understood that the geographic labels used here do not refer to geographical directions. We simply borrow the idea of geographical directions for our labels because it is simple and familiar, and it is also more economical (in an exposition such as this) to say that a ball is traveling 'west,' than to say it is traveling 'in a positive direction along the x-axis.' Unlike geographical directions, we can orient our three-dimension scaffold in any arbitrary manner for our convenience. We ask you to understand the three axis's we have labeled with directions is the same as that labeled x, y, and z or labeled anything else. The only requirement is that the three axis's are kept at right angles.

Now to say that motion is conserved in every single direction is logically equivalent to the idea that motion is conserved in each of the three dimensions, or in each of the six directions. This immensely reduces the difficulty of computation in our three dimensional world. We shall return to this idea. We should add another idea to this preamble. Contained implicitly in the notion of direction is another fundamental notion, that of 'straight.' You can travel between two points, say points A and B, by means of an infinite number of different paths, but only one of those paths can be uniquely described as a 'straight line' or 'straight path.' It is also the shortest path. We cannot describe the 'speed' or 'velocity' of an object without implicit use of the notion of the straight line. In physics, as in geometry, it is fundamental. And it is also axiomatic in that it cannot be defined in terms of other concepts. (The concept of 'straight' always re-appears implicitly in such "definitions," for instance, in the notion of 'shortest.') When we say that a object travels a certain distance, even along a curve, we can only define (and measure) the notion of 'distance' by utilizing the concept of 'straight.' As a consequence, to measure a quantity-of-motion requires the underlying idea that the motion is in 'a straight line,' and to say that a quantity-of-motion continues in a 'straight line' at a constant speed, unless other motion is added, is taken here to be axiomatic.

The above stated 'axiom' (essentially Newton's first law of motion), must be qualified only in regard to the particle-points. Although we must postulate that 'freely moving' particle points do move in straight lines, we must also postulate they are can be aggregated at the most fundamental level, that is, 'hooked' together in some manner. This is a complicated subject which we will return to later. According to Newton, any time a mass (or in our scheme, a quantity-of-motion) is 'acted upon' by a 'force,' an acceleration occurs and motion is exchanged. Can this formulation be stood on its head? Can we say that anytime motion is exchanged, the exchange can describe as a 'force.' A force, in this interpretation, simply demarcates or categorizes certain types of exchanges. This idea will likely violate most people's notion of a force as something which 'causes' motion or causes alterations in motion. However, if the concept of 'force' (with a small 'f') is understood as a shorthand description of a principle or law of exchange, we need postulate no mysterious 'agency.' Physicists can in fact describe the motions and the exchanges precisely and in great detail, but the so-called force (as an agent) cannot be described or explained in any manner whatsoever except as 'that which causes motions.' In this regard, it seems of questionable value. Generally speaking, the task of explanation is to reduce (explain) concepts in terms of other more fundamental concepts, until at last we arrive at the most fundamental or 'ultimate' concepts which must be accepted as irreducible. We are inclined to regard these fundamental concepts as 'self-evident.' Of course, controversy has always raged about what is or is not 'self-evident,' and most likely will continue to rage. Yet over the long millenniums of human investigation, certain concepts have been slowly winnowed out, such as the principles of non-contradiction, the principle of sufficient reason, and a host of others which are the basic presuppositions underlying present day scientific and mathematical investigation. They are widely accepted but not sacred. (So perhaps we should add to them, the 'principle of fallibility,' a cautionary heuristic, which as Descartes children, we must honor above all others.)

We will return this subject momentarily, but let us first return to our billiard table, a real world table, ignoring only the effects of friction and inelasticity. The cue ball is again struck and accelerated to a speed of ten meters per second. We make careful note of its direction. Assume that it travels straight down the center of the table in a direction we label 'due north.' It soon strikes another stationary ball slightly off-center. In this particular real-world scenario, assume the cue ball moves off on a north-westerly course at 6 meters per second, and the struck ball moves off on a north-easterly course at 8 meters per second.

Now it is readily apparent that the sum of the speeds of the two balls subsequent to the collision is not 10 (as it was in our fanciful world) but 14. Obviously, the "sum of the speeds" of these balls is not conserved in the manner that it was in our fanciful world. The speed of the cue ball in the real world is somehow transmuted into a combination of speeds after the collision that has a value greater than 10.

Nevertheless, we can perform certain observations. Given the frame of reference that we have imposed, we notice that the ball traveling northwest can be said to be traveling in two directions simultaneously, the direction of 'north,' and the direction of 'west.' In other words, the direction of the balls can be 'decomposed' into a combination of the fundamental directions. The cue ball began its journey by traveling 'due north' at a speed of 10 meters a second, which is to say, it traversed 10 meters on the billiard table in one second. (Assume the billiard table is very large.) After the collision it changed course and traveled northwest a distance of 6 meters in each second. However, we can also measure the distance that this ball travels in the strictly north direction in each second and the distance it travels in the strictly west direction in each second, each measured separately. We find that it travels a distance of 3.6 meters north and a distance of 4.8 meters west. We have separated the direction "northwest" into its two directional components, the north and the west. Physicists refer to these components as 'vectors.' (You can confirm these measurements with the Pythagorean Theorem: the sum of the squares of 3.6 and 4.8 is equivalent to the square of 6, the hypotenuse.)

Now if we measure the motion of the other ball (the stuck ball) which is traveling with a speed of 8 meters per second in a northeasterly direction, we find that it is traveling with a speed of 6.4 meters per second in the north direction and 4.8 meters per second in the east direction.

Certain observations can now be made concerning this idealized real world experiment. If the two north components (or vectors) of the two balls (3.6 and 6.4) are summed, the result is 10. The two components sum to the same value as the speed of the cue ball. We cannot say that the components represent the speeds of the balls, only that they represent the speeds in the *north direction*. We can say that the sum of the speeds of all the north-vectors is equivalent to the speed of the originating cue ball. In other words, the *sum of the speeds is directionally conserved.*

Notice also that the speed in the west direction is equivalent to the speed in the east direction, both being 4.8. If we perform the 'trick' of assigning a negative number to one of these directions, they will sum to zero. This is to say, that the *net* speed in the east and west direction of the entire assemblage after collision is exactly equivalent to that before the collision, namely zero.

These findings will hold true no matter what the masses of the balls, so long as we multiply mass time speed to find what we have termed a 'quantity-of-motion.' That quantity is conserved, and it is known as the conservation of momentum. We can refer to it as the conservation of a 'quantity of motion' *directionally*, which is to say that the *net motion* in each direction has not changed after the collision.

Another observation can make this idea more compelling. Suppose the collision involved a cue ball and a cluster of nine other balls at the center of the table (as in a cue-off). The nine balls clustered together (combined with the cue ball at the moment of collision) have what is called a 'center of mass.' The concept of 'center of mass' can be easily understood with the help of a simple illustration. Suppose you pack a box full of china somewhat carelessly. The box is not evenly packed but heavier on one end than on the other, heavier on one side than the other, and heavier at the top than at the bottom. Since weight is proportional to mass, you can easily find the 'center of mass' of the packed box by placing the box on 'knife edge' parallel to one of its sides. The box will not balance at the center but tip toward the heavier side. You move the box until it balances. Then you extend an imaginary plane through the box, as if the knife edge were to slice it in two. Now you rotate the box ninety degrees and perform the same operation. Then tip the box on one of its ends, perform the same operation again, and extend another imaginary plane through the box. These three planes will intersect within the box at a certain point, the center of mass. This point may not correspond with any object in the box, but it represents the point at which the mass is equally placed about this center point.

Every object, large or small, has a 'center of mass.' And in a like manner, the cluster of balls has a center of mass. The cluster will have a center of mass even *after* it is struck by the cue ball. Let us say that after the cluster is stuck by the cue ball (traveling at 10 meters a second due north), the balls (including the cue ball) fly off in all directions on the table. Some may even rebound southward. Yet if all the resulting motions are summed, those vectors in the east direction will equal (and exactly cancel) those in the west direction. And the vectors in the south direction when subtracted from those in the north, will leave a net north motion of 10, exactly the original motion (speed) of the cue ball. And if we locate the center of mass of this spreading cluster of balls, we can observe that it will be traveling due north at a speed of 1 meter per second. The original motion of ten meters a second for the single cue ball has transmuted to the final motion 1 meter per second for the mass of ten balls (the 9 in the cluster plus the cue ball) for a total motion of 10 units. In other words, if we could 'freeze' the motions of the cluster of balls as they spread outward on the table at each succeeding second, and compute the center of mass, that center would move just as described. It would act as if it were a single spreading object with a mass of 10 which, receiving the motion of the cue ball, moved north exactly as would a single ball of mass 10 (the 9 balls in the cluster plus the cue ball). If we are allowed to think of the spreading cluster of balls not as an array of objects, but as a single object spreading outward or growing in size, its 'essential' motion (that of the center of mass) would exactly mimic that of a single (none spreading) object. And if that is the case, we can conclude that motion is in fact conserved directionally. A certain amount of motion north will always remain a certain amount of motion north whatever the permutations and transmutations that may occur within the particles of the larger whole.

Perhaps you will object at this point. This is all well and good, you might say, but surely some additional motion has mysteriously appeared. Notice that in the case of the collision of the cue ball with the stationary ball, two vectors (east and west, both of which were 4.8) were subtracted to sum to zero. But can they not also be *added* to sum to 9.6? Surely the ball traveling west involves a certain quantity of motion, and the ball traveling east involves a certain quantity of motion. Subtracting these motions enables us realize that the *net* quantity of motion is conserved, but that should not obscure the fact that these two vectors, considered individually, both represent a quantity of motion which must be accounted for. How can the quantity of motion (as we have so far defined it) be conserved if there is an additional sum of 9.6?

Nature has her way and her reasons. We can begin to understand this if we first examine a fatal shortcoming of our fantasy universe. Like all such fantasy universes it is unworkable. It is easy to create superficial, unworkable fantasy universes, but to create a universe that is logically consistent and that will account for the diverse phenomena of the real universe is no easy matter. The 'rules' which govern our real universe are being slowly unraveled by investigations beginning a few millennia ago. We still have far to go. Some researchers believe that there is only one set of rules that will create a real universe and that no other workable set exists or could exist, but we need not delve into that proposition here. However, a working presupposition for all physicists is that a set of rules exists for the real world. It is their job to find them, similar to what mathematicians do. Mathematicians and physicist both investigate the same universe but different landscapes.

As for our fantasy universe, consider this case: a one kilogram billiard ball traveling at 10 meters a second collides 'head on' with a 5 kilogram stationary ball. Since the collision was not at an angle, the two balls will travel after the collision along the same axis. Suppose the first ball conveys all it motion to the second ball. The second ball would then have to travel at 2 meters per second. A 5 kilogram ball traveling at 2 meters per second has the equivalent motion of a single ball one kilogram ball traveling at 10 meters per second.

But the question which arises is not what will happen *if* all the motion is transferred, but more simply: what will happen? For there seems to be quite a few possible outcomes. For instance, if after the collision the 5 kilogram ball moves at 1.8 meters a second (1.8x5 = 9) and the one kilogram ball travels in the same direction at one meter a second (1x1= 1), the sum will also be 10. Or if the 5 kilogram ball travels at 1.9 meters a second (1.9x5= 9.5) and the other ball at .5 meters a second (1x.5=.5), the sum will again equal 10. It turns out that there are an infinite number of possible combinations of speeds which will sum to 10. So what will be the actual outcome? Nature cannot allow the billiard balls to freely decide this matter for themselves. Nature is *determined,* a fundamental presupposition of all scientific enquiry whether or not you subscribe to it as a metaphysical proposition. It turns out, somewhat counter-intuitively, that the 1 kilogram ball will *rebound* and travel in the opposite direction of the 5 kilogram ball.

In our fantasy world we can state the dilemma with this equation:

$$(mass_1)(speed_0) = (mass_1)(speed_1) + (mass_2)(speed_2)$$

or more succinctly:

$$m_1 s_0 = m_1 s_1 + m_2 s_2$$

or: $1 \times 10 = 1 \times s_1 + 5 \times s_2$

where $mass_1$ is the one kilogram ball and $speed_0$ is the original speed of 10 m/s, and $mass_2$ is the 5 kilogram ball. $Speed_1$ and $speed_2$ are the two unknown speeds after collision. But this one equation has two unknowns. It cannot be solved for a unique solution. Something more is needed to determine the outcome.

It is evident that the conservation of directional motion (momentum) imposes a constraint on the motion of the balls. If motion is *directionally* conserved, the additional motion to the east and west must be equivalent in order for them to cancel out. This imposes a constraint absent in the unreal fantasy world discussed earlier. We can formulate this constraint with the following simple formula:

$$m_1 v_0 = m_1 v_1 + m_2 v_2$$

where m_1 is one (the mass of the first ball), v_0 is 10 (the initial velocity), and m_2 is 5 (the mass of the second ball). The values of v_1 and v_2 are the two unknowns. We term them 'velocities' and not 'speeds' because we have incorporated the directional constraint. Yet, as illustrated before, this single equation cannot be solved to yield a unique value for these final velocities. In other words the single constraint is not sufficient to determine the outcome.

So let us look again at the numbers involved in the real world case. The cue ball had an original speed of 10. Subsequent to the collision, the speeds of the two balls were 8 and 6. There is relationship between these three numbers which is not difficult to spot. The square of 6 is 36, and the square of 8 is 64. The square of 10 is 100. The sum of 36 and 64 is 100. The *squares* of the speeds of the two balls after collision is equal the square of the original speed.

$$m_1 v_0^2 = m_1 v_1^2 + m_2 v_2^2$$

In this peculiar way, the sum of the speeds *is conserved*. It is conserved as the sum of the squares. This is a second 'constraint' placed upon possible motion. The first constraint was that the motion must be conserved directionally; the second constraint is that it must be conserved as a sum of the square of the speeds. We now have two equations for the two unknowns which can be solved for a unique real value.

$$m_1v_0 = m_1v_1 + m_2v_2$$
$$m_1v_0^2 = m_1v_1^2 + m_2v_2^2$$

When the one kilogram ball strikes the 5 kilogram ball (under the conditions prescribed), we find that the final velocities are: v_1 = -6.7 m/s, and v_2 = 3.3 m/s. In other words, the 1 kilogram ball will rebound.

The sum of the velocities squared (equivalent to the sum of the speeds squared) provides us with a second conservation rule which allows motion to be uniquely determined. Yet it seems intuitively less obvious than the sum of the speeds proposed in our fanciful universe. Why do the speeds need to be squared and added, rather than just added? What is the intuitive basis for this necessity?

There are two answers. For one thing, if there existed *only* directional conservation of motion, that alone would be unable to generate our universe. If the initial motion of the cue ball (due north) was conserved *without* the additional motions (east and west), all collisions could only result in subsequent motion in the original line of travel. This is a problem recognized (in a sense) by Democritus and Lucretius two thousand years ago. Democritus proposed that at the beginning of time, atoms were "raining" straight downward in a great homogenous mass. You might imagine a rain storm on a calm day. Since the atoms were dropping vertically and uniformly, they did not touch. And since they did not touch, they could not combine. Nothing could come of such a world. So Lucretius introduced a 'swerve.' In the course of the fall, an atom that swerved sideways could collide with another atom. This would upset the equilibrium and result in a concatenation of collisions from which all objects of the world would be assembled. The collisions produced material objects because the atoms became 'hooked' together. These early day theorists did not explain the origin of the 'swerve,' nor were they very clear about the nature of the 'hooks.' We can forgive them for the problem of the 'hooks' because science still struggles with it.

A certain view of modern cosmology has some parallels with this vision of Democritus. In this view, there was in the beginning a great rainstorm call the Big Bang. The "Atoms" – which is to say, the initial constituents of matter - rained 'outward' in equilibrium in a great homogenous flux. This equilibrium was then quickly broken by a "swerve-like" dis-equilibrium called a 'quantum fluctuation' which resulted in the condensing and assembly of the initial constituents. What allowed the constituents to coalesce into macro objects were "hooks" – now called 'forces.' The forces are said to 'act' upon the constituents of matter, holding them together in a hierarchy of ever larger macro objects.

For present purposes, we suggest that "swerve" (or its functional equivalent) is provided by the two principles of the conservation of motion working in tandem. The directional conservation of motion provides order. This means that the particles (whatever they may be: billiard balls or atoms or anything else) cannot travel in any old way. If they could travel in any old way, the universe would be unordered, chaotic, unpredictable and unintelligible. Instead they are constrained to travel in a straight line in a definite direction (other things being equal).

Order and predictability is established. Nevertheless, if they could *only* travel in a single direction (raining straight "downward" or "outward"), nothing could assemble. Orderable deviation from the straight directional path is required (the 'swerve') if particles are to combine. The order is provided by the principle of the conservation of the sum of the velocities squared – the 'total' motion.

Before we delve a bit further into this explanation, let us examine some simple equations of motion which will help set the stage. We begin with the simple equation for velocity which is given by the formula:

$$v_f = at$$

Here, v_f is velocity, a is acceleration, and t is time. Our formula says that velocity is equal to the rate of acceleration multiplied by the time over which the acceleration occurs (assuming uniform acceleration). For instance, if a body is accelerating uniformly at 10 meters per second per second for 5 seconds, then the final velocity obtained will be 10x5 = 50 meters per second. This seems physically as well as mathematically almost self evident. It can be empirically confirmed. For instance, if an object is dropping (and accelerating) at 10 meters a second per second near the earth's surface, it will obtain a final velocity of 50 meters per second after 5 seconds.

It seems odd to say that the formula can be empirically confirmed. Let us say the object drops for 5 seconds and accelerates at exactly 10 meter/s^2. Now let us say that physicists measure the velocity at the terminus of the fall with the most accurate instruments available, and determine that the final velocity is not 50 m/s but, say, 60 m/s. This would be quite stupefying, not only to physicists, but to mathematicians as well. They would both passionately insist that some measurement was wrong, not that the formula $v_f = at$ is wrong. Does this mean that nature 'obeys' mathematics? Or should we ask: does mathematics 'obey' nature? Our answer is that the two cannot be separated. They are two aspects of the same reality.

But let us return to our formulas. If velocity is equal to acceleration multiplied by time ($v_f = at$), then we can rearrange terms to give:

$$t = v_f / a$$

Here, time is equal to velocity divided by acceleration. For instance, if a body is traveling at 50 meters a second and has been accelerated from rest at 10 m/s^2, then the time of acceleration can be computed as 5 seconds.

A simple formula for the distance traveled by such a body is given by:

$$d = v_a t$$

Here, d is the distance. This formula says the distance traveled is equivalent to the velocity multiplied by the time. We have placed the subscript (a) to indicate that this is an average velocity, to distinguish it from v_f where the subscript (f) indicates final velocity. In other words, if a body is accelerated to a final velocity of 50 meters a second from rest, the average velocity over this period of travel will be 25 meters per second, one half of the final velocity, assuming uniform acceleration. The distance traveled will be 125 meters if the time of travel is 5 seconds.

Now we can substitute $t = v_f / a$ into the equation for distance to give:

$$d = v_a v_f / a$$

Rearranging these terms gives the odd looking formula:

$$ad = v_a v_f$$

Here, acceleration multiplied by distance equals the average velocity multiplied by the final velocity. Since v_a is equal to one half v_f , we can re-write the formula:

$$ad = ½\, v_f^2$$

We shall return to this formula in a moment. There is another simple formula for distance, only slightly more complicated, which will give us the same answer:

$$d = ½\, at^2$$

In this case, the distance traveled for a body accelerated from rest is equivalent to the acceleration multiplied by the time squared. For instance, if a body accelerates at 10 m/s^2, over a period of 5 seconds, the distance traveled will be 125 meters (as above). This formula can be easier understood if it is written:

$$d = ½\, (at)t$$

In this case, it is apparent that (at) equals the final velocity v_f (from above). Therefore $d = (½\, v_f)t$. Since one half of the final velocity ($½\, v_f$) is equal to the average velocity, v_a , (assuming uniform acceleration) the formula reduces to $d = v_a t$, as above.

If we substitute $t = v_f / a$ (which is time equals final velocity divided by acceleration – from above) into $d = ½\, at^2$ we have:

$$d = ½\, a\, (v_f /a)^2 \quad \text{or} \quad d = ½\, v_f^{\,2} / a$$

Again, rearranging terms gives the odd looking formula derived above:

$$ad = ½\, v_f^2$$

If we take the liberty to multiply each side of this equation by a constant - what we have called ‘units of motion’ or units of mass, designated by the letter ‘m’ - then we have:

$$(ma)d = ½\, mv_f^2$$

Physicists refer to the term on the left as ‘Work’:

$$W= (ma)d$$

They refer to the term on the right as ‘Kinetic Energy’:

$KE = ½\, mv_f^2$

In our scheme, the number given by these terms represents a 'quantity of motion'. In the first case, it is calculated from acceleration and distance and is called Work; in the second case, it is calculated from the final velocity and is called Kinetic Energy.

We have already interpreted mass multiplied by acceleration *(ma)* as Force - the rate at which motion is transferred. So we can interpret the formula for Work in this way: if motion is transferred at a certain rate *(ma)* over a certain distant *(d)*, the resulting quantity-of-motion transferred can be calculated with the formula: *(ma)d*. This quantity can also be calculated as one half of the square of the final velocity obtained over that distance $(½\, mv_f^2\,)$. The final velocity is obtained by accelerating the mass at the given rate '*a*' over that distance.

For instance, suppose a body of 1 kg is accelerated over a distance of 125 meters at 10 m/s^2. Calculation gives the quantity-of-motion *[(ma)d*] as 1250 $kg\text{-}m^2/s^2$. The time required for travel will be 5 seconds (from $d = ½\, at^2$), and the final velocity of the body will be 50 m/s. Consequently, the same 'quantity of motion' can be calculated using the final velocity (50m/s) by the formula $KE = ½mv_f^2$, which equals ½ (1) $(50)^2$ or 1250 $kg\text{-}m^2/s^2$.

Now let us return to the second answer to the question asked above: why should nature use the sum of the velocities *squared* to represent this manifestation of total energy rather than something 'simpler' and more intuitive, like the sum the speeds? The sum of the speeds *squared* seems intuitively less obvious than the simple addition of speeds. Yet Nature has a second reason which is as much mathematical as it is physical (although we make no sharp distinction between the two realms, nor do we believe does Nature). Let us return to our imaginary billiard table for a moment and imagine this situation: a billiard ball is traveling 'due north' down the center of the table at 8 m/s. Now imagine that a second identical ball is traveling perpendicular to this ball, in a direction due west at a speed of 6 m/s. Imagine that the second ball collides with the first and transfers its entire motion to the first ball. In this *idealized* situation, it is now apparent that the first ball will have acquired a second component of motion (6 m/s west) and will consequently be comprised of two components – one north and one west. It has the original component north of 8 m/s, but also the additional component west of 6 m/s contributed by the second ball (which comes to a stop). Since motion is *directionally* conserved, the ball must now travel 8 m/s in the north direction and 6 m/s in the west direction. It now 'contains' both components and must travel accordingly. If it did not travel in this fashion, motion would be lost. So the ball will consequently continue to travel 8 meters in the north direction in each second – and at the same time, 6 meters in the west direction in each second. The path will be northwesterly, a diagonal with respect to the frame of reference of the table. To calculate the speed of the ball, we must find the length of the path traveled in each second. This can be done with the Pythagorean Theorem: $(\text{distance})^2 = 8^2 + 6^2$. Consequently, the distance actually traveled in each second is 10 meters, not 14 meters (the sum of 8 + 6), and the speed is obviously 10 meters per second.

Accordingly, it can be said that the ball traveling in the northwesterly directions 'contains' – in a manner of speaking – the two component speeds, or any other component speeds whose squares sum to 100. This is not merely an empirical relationship but a mathematical necessity. In other words, the directional conservation of motion logically entails the conservation of the square of the component velocities. If the 8 meters-per-second motion is to be conserved, and if the 6 meters-per-second motion at right angles to it is also to be conserved, the resulting motion 'contained in' the traveling one kilogram mass will require that the mass travel northwest at 10 meters a second. This is to say that the squares of the velocities must be conserved if motion can be exchanged and combined.

To see this more clearly, imagine that the ball traveling in the northwesterly direction at 10 m/s strikes another ball off center such that the struck ball travels due west at 6 m/s, and the first ball travels onward in a north direction at 8 m/s. (This is the first collision analyzed above.) Here, the original motion of the two balls has returned. As long as the principle of directional conservation of motion is maintained, we can derive the principle that the sum of the speeds of all masses in motion can be calculated as the sum of the speeds squared (or equivalently, velocities squared) and this sum will also be conserved. So the total motion – that is, the sum of the speeds of all identical masses in any isolated system will be proportional to the sum of the speeds squared. More generally the sum of the masses multiplied by the speeds squared (where the masses are unequal) will also be conserved. (This is true so long as we conceive of masses as 'units of motion' as described earlier.) The relationship is expressed below:

$$m_1 v_0^2 = m_1 v_1^2 + m_2 v_2^2$$

where m_1 and m_2 are the two masses, v_0 the initial velocity, and v_1 and v_2 the resultant velocities.

The question that remains is why is Kinetic Energy given by the formula KE = ½ mv^2, not KE = mv^2? Why the ½ ? The answer can be found is the nature of the exchange of motion. Whenever motion is exchanged between two bodies, the motion is not exchanged instantaneously. Instead, the two bodies will accelerate. The body receiving the motion will speed up relative to a given frame of reference, and the body giving up the motion will slow down. The exchange takes place "over time," not "in an instant." This introduces an average velocity of ½ times the final velocity into the equation.

The question of why the exchange does not occur instantaneously is interesting and complicated. It is not something most people would question. It seems odd to ask why you cannot get into your car, turn on the key, and instantaneously transfer to 60 miles per hour. Such a thing is completely foreign to our experience.

Isaac Newton postulated that the moon was held in orbit by a force acting instantaneously 'at a distance.' The notion of 'action at a distance' was much ridiculed by Newton's contemporaries, for it seems obvious that you cannot push a cart by walking 10 meters behind it - exerting "action at a distance.' Nor can you pull the cart by walking 10 meters in front of it - exerting 'action at a distance.' In the first case you need to put your shoulder to the cart, and in the second case you need a harness. The notion of action-at-a-distance seemed far-fetched. Newton was not happy with it himself, but he could think of nothing better. As a good practicing physicist, he decided to simply accept it and move on.

The "far-fetched" notion of 'action at a distance' has a curious history. For it was soon pointed out that magnets seem to affect iron particles 'at a distance,' and electrostatic forces also seem to act 'at a distance.' The paradigm changed. 'Action at a distance' no longer seemed far-fetched. It appeared that all forces 'act at a distance,' at least if the distance is very small. For instance, the distance between molecules in a substance is small, but the molecules must still be held together by forces 'acting at a distance,' or so it seemed. 'Action at a distance' no longer seemed preposterous; it became respectable, even "self-evident."

The view changed again with the advent of quantum mechanics. The theory called the 'standard model' postulated that forces are 'carried' between objects by special 'carrier particles.' The carrier particle for the electro-magnetic force is the photon. The carrier particle for the gravitational force is a postulated 'graviton' (which so far remains undiscovered). The idea that forces are 'carried' by particles may seem as peculiar as action-at-a-distance. Of course, if a billiard ball is struck by the cue stick and rolls across the table to strike another billiard ball, it could be said that the 'force' (in some loose sense) is carried by the ball as it rolls across the table. The rolling ball is obviously the mechanism by which the energy (or motion) is transferred across the table from the cue stick to the second ball. On collision, however, photons deliver the force to the struck ball. How this occurs, in detail, is problematic. You might ask: what carries the force from the photons to whatever they interact with?

Is there a sub-photon which is emitted by the photon at very close range to 'carry' the force from the photon to the nucleus? This idea perhaps seems far fetched, but obviously the model of 'force transfer' cannot continue with ever smaller sub-particles. If it does not continue in this way, at some point the particle must transfer the force. The force will have to be transferred 'at a distance' (however small that distance) - or not. If the force acts at a distance, it would have to act instantaneously (at a distance) if there were no intermediaries to carry the force the small distance from particle to particle. (If the force did not act instantaneously, any delay would call for an explanation.) Instantaneous action at a distance will have reappeared, albeit at a small scale. On the other hand, if the force does not 'act at a distance,' why is it even called a 'force?'

Let me add some explanation here. Setting aside anthropomorphic notions of agency, we usually conceive of a force as that which 'causes' something to move in the physical world. In this sense, it is an invisible connective. If you have two magnets which seem to affect each other 'at a distance,' we can explain that by saying that there is something 'there' pushing or pulling, namely, the 'force.' The forces are usually connected with something called a 'field' or a 'force-field'. Fields are generally conceived of as 'something' which surrounds the object emitting the field. The field will explain (or describe) how forces will act, most notably in which direction. The forces are the things which do the 'acting' and which cause motions. As long as the forces are acting 'at a distance' this model seems to make some sense, for there must be a 'something' to connect the two bodies, the acted upon and the acting. But if the two bodies are no longer separated by a distance, the question arises as to what exactly do forces do?

We do not see the forces, we see only the effects (or the alleged effects) of the forces. We see and can study and measure the motions resulting from the alleged forces, but not the forces themselves. We measure the 'strength and nature' of the forces only by cataloging the effects. As was mentioned earlier, it could be said that a "force" is nothing but a handy short hand description of the motions which occur (or which are liable to occur) when bodies are in proximity to other bodies and when motions are exchanged. In this sense, bodies of a certain nature have a propensity to accelerate in proximity to other bodies, and these propensities can be 'mapped."

The difficulty is that forces seem to have no 'intrinsic properties.' They have only 'properties of effect.' The difference between intrinsic properties and properties of effect can illustrated with the concept of 'wind'. Wind obviously has properties of effect. It can billow a sail and drive a ship across the ocean. But wind has intrinsic properties as well. It consists of moving air (atoms of nitrogen and oxygen) and host of other 'inherent' properties. We can measure and study these properties in great detail. The alleged ‘forces,’ on the other hand, have no intrinsic properties that we know of. If you explain that things move because of a force, and then define the force as ‘that which causes something to move,’ the narrow circularity of this definition makes it seems as if the force has been defined into existence.

If I may digress, there were other alleged entities which were proposed to account for effects. “Phlogiston” was a theoretical ‘substance’ conceived in the seventeenth century to explain combustion. If you put a match to a piece of wood, it will burn, but if you put a match to a stone, it will not. Why the difference? This is a perfectly good question and the answer proposed was that the wood had ‘phlogiston’ and the stone did not. Phlogiston was the thing responsible for the combustion of the wood. Why did the wood burn: because it had phlogiston. What is phlogiston? Phlogiston is ‘that which causes the wood to burn’! You can see the obvious similarly between the concept of ‘phlogiston’ and the concepts of ‘force’ and ‘field.’ It might appear that in both cases, something has been ‘defined into being.’ Of course, phlogiston was not imagined in that way. Rather, it was imagined to be a ‘substance’ of some kind, an ‘air’ or ether which inhered in the wood but not in the stone. (Perhaps much in the way that some theorists might imagine ‘forces’ or ‘fields.’) But no one could discover any intrinsic properties of this alleged substance. It’s only definite property was the alleged property of effect.

Another interesting theory proposed in the eighteenth century was the concept of the ‘caloric fluid’ used to explain the phenomenon of heat. It was observed that heat seems to flow from hot substances to cold ones. Was this not due to a substance, a fluid, which flowed from the one to the other? Thus was born the concept of ‘caloric fluid.’ It seemed to be a satisfactory idea, except that the inherent nature of the fluid was unknown.

In the end, it turned out that this theory was substantially correct. Heat was indeed a fluid that flowed from the hot body to the cold ones. However, the fluid was not ‘caloric’ but a fluid called kinetic energy. One mysterious fluid (energy) was substituted for another (caloric). Count Rumford observed that the heat-fluid kept pouring out of canons as they were being bored so long as the boring continued. It was just a short step to see that the fluid was pouring into the canon (by the friction of the boring mechanism) as well as out.

The concepts of ‘phlogiston’ and ‘caloric fluid’ were steps in the right direction. If matter is conceived in Aristotelian terms as ‘earth, air, fire, and water,’ then it would seem that the wood must contain some fire, but not the stone. The concept of phlogiston pointed the way to the recognition that wood has inherent properties responsible for easy combustibility that stone lacks. When it was understood that fire is form of rapid oxidation, the theory of phlogiston was simply discarded, but it clearly pointed towards a theory matter at odds with Aristotle’s. And the theory of a caloric heat fluid, though wrong in detail, was right is essence. Incorrect theories do not necessarily lead in the wrong direction. (We hope you will keep this in mind regarding our efforts here.)

None of this need imply the concept of ‘force’ is illegitimate; only that it should be suspect. If you are philosophically inclined, you might argue that at some ultimate level you must find concepts which cannot be ‘reduced’ (explained) in terms of other concepts. These fundamental concepts are ‘irreducible.’ So we again arrive at the question as to whether force can be explained in terms of other concepts and principles of a more fundamental nature. Often our satisfaction or dissatisfaction with a concept springs from our intuition that it has or does not have that taste of the ultimate

This philosophical digression, accompanied by more questions than answers, has caused us to stray from the subject at hand: why is kinetic energy measured as ½ the mass times the velocity squared. The answer is closely tied to the exchange of motion. If motion was exchanged instantaneously, nothing would exist, and time would cease. Instead, motion 'flows' from one object to another over time. Since motion is not exchanged instantaneously at anything other than possibly the quantum level, the exchange is accompanied by accelerations. Since the quantity exchanged is a function of velocity, and since the velocity continuously changes over the period of acceleration, the quantity must be measured as proportional to the average velocity and not the final velocity of the accelerated object. The average change in speed under uniform acceleration will be one half the final velocity (if accelerated from rest), or the difference between the initial velocity (if the object is initially in motion) and the final velocity divided by two. This relationship can be expressed as the integral or 'summation' of the momentum between two velocities. In the calculus, the integral is written as:

$$m\int v = m[1/2v^2 - 1/2v_0^2] \quad \text{(or } 1/2mv^2 \text{ where } v_0 \text{ is zero)}$$

We must concede that the idea of 'velocity squared' has no intuitive counterpart in ordinary experience. The idea is expressed in the dimensions of meters-squared divided by seconds-squared (m^2/s^2). This can be contrasted to meters-per-second (m/s) which is speed, a familiar concept; and meters-per-second-per-second (m/s^2) which is acceleration, another familiar concept. According to the scheme proposed here, velocity squared is a mathematical relationship which can best be understood as a 'quantity of motion.'

Of particular note is the fact that the kinetic energy increases as the square of the difference in velocity. For instance, the kinetic energy of a one kilogram ball traveling at 10 meters per second is four times as great as the kinetic energy of the ball traveling half as fast at 5 meters per second; (100 kg-m^2/s^2 versus 25 kg-m^2/s^2). Yet the total absolute value of the momentum after collision will sum to a value directly proportional to the difference in the speeds, not, as with energy, to the difference squared. In other words, let us say a ball traveling at 5 m/s collides with another identical ball. Suppose the energy is equally divided. The two balls will each travel after collision at 3.54 m/s. Now a ball traveling at 10 m/s (which has four times as much energy) which undergoes a similar collision will result in the two balls traveling at 7.07 m/s each. The speeds in second case are twice as great as in the first case, not four times as great. There is no mysterious increase in the resulting speeds of the vectors. We can say with all due justice, that energy (as sum of 'velocities squared') is simply nature's mathematical way of conserving the value of the speeds.

Now we shall leap forward over a great deal of intermediary material to arrive at an elucidation of Einstein's famous formula. We will then step back and fill in some of the large gaps left behind and explore some of consequences of this proposition.

We ask you to imagine a single stationary billiard ball suspended at the center of a large imaginary 'box.' Do not worry about the nature of the box. The box represents a hypothetical isolated region of space. Now imagine that the ball inside the box can be made to "dissolve" into all its constituent parts. Suppose the 'bonds' holding the constituents parts together can be magically undone. We do not suggest that this is an actual possibility, but only an imaginative possibility. Suppose the molecules fly apart and dissolve into atoms which dissolve into their constituents: the protons, neutrons, electrons, and all other constituents. These constituents in turn dissolve into their constituents. The gluons, photons, and all particles of radiation dissolve along with the "strings" (if there are such things). What are we left with?

We suggest that in this case the end result will a great rapidly diverging gas of particle-points. They are the ultimate constituents of nature. They will radiate outward from the center of this small mass, equally in all directions. The mass will have been "dissolved" into its most fundamental elements. Total net momentum will be zero with respect to the box because if it was not, the ball would not have been at rest. If there was any net motion in the assembly of the constituents, the ball would have been moving in that direction. Since it was at rest, the net motion is zero relative to the box.

The number of particle-points will be unfathomably large, but finite. They will be inherently identical in nature. We assume this to be true from the principle of parsimony and the principle of hierarchy which will be discussed later. This is to say that each of these particle-points will travel in their free "unbound" state in a straight line with the same identical speed, the speed of light (c). Each particle-point assumes its identity in being a separate and individual point in motion traveling at the same identical speed.

Of course, the *manner* in which they move (other than their inherent speed) is the subject of physical inquiry. The general principles of the quantification of motion, which have been the subject of this inquiry, must devolve to 'the general principles of the aggregation of motion' and 'the principles of the exchange of motion.' These principles are not entirely separate and they form the primary subject matter of physical enquiry.

The total motion inside the 'box' can be summed. Since each particle-point travels at exactly the same speed, we need only sum up (or count) the number of particle-points in the box and multiply the sum by the speed squared (c^2). Of course, counting is not practicable; we must rely instead on calculation, but here we arrive at the ultimate unity of mass and energy: for each unit of mass (each particle-point) is a unit of energy, and each unit of energy (each particle-point) is a unit of mass. In other words, what we term "mass" is a proxy for the number of particle-points contained in the object whose mass has been determined by other means. To compute *energy* – the quantity of motion in an aggregation of particle-points, we need only multiply that number (the mass) by c^2. In this manner we arrive at the fundamental equation for the total energy (or 'quantity of motion') in any quantity of mass:

$$E = mc^2$$

The appearance of this peculiar constant (c^2) is consequently closely related to the familiar equation for kinetic energy. The reason why it is c^2 and not ½ c^2 in this special case is because particle-points do not undergo accelerations. Accelerations result from the exchange of motion – or in the most fundamental terms, the exchange of quantities of particle-points. But particle-points cannot themselves be accelerated, an idea that has far reaching consequences which we will look into presently.

Each fundament increment of motion is thus an increment of mass, and each increment of mass in an increment of motion. Since these increments of motion have an identical speed, it is evident that the *constant* c^2 serves to specify the speed squared. Unlike the *variable* v^2 used in the computation of the energy of quantities-of-motion-in-motion, c^2 is unchanging. On the other hand, when a quantity-of-motion (a mass) is accelerated, increments of motion are added to it, and the velocity is increased as does variable mass (m) (relative to a stationary framework).

The particle-point with its sole property of speed is the ultimate Democritean atom, but in a Hericlitean universe. Just as there is a certain number of molecules in a kilogram of water, or a certain number of atoms in a kilogram of hydrogen; there is a certain number of particle points in a kilogram of all substances. The analogy holds.

If motion is the ultimate 'substance' of our universe, then it is apparent why things move – to answer the question posed earlier – they move because motion is their essential nature.   The inherent motion of the particle-points is the ultimate 'something' that drives all things to their destinies

Yet could such a spare conception possibly account, in some ultimate reckoning, for all things in our universe, all that has ever been and will ever be – from the hydrogen atom to the Mona Lisa? Can these simple objects, the pawns of nature, truly be assembled into the extraordinary variety of all that is or will be? This is a daunting idea. However, it must be stressed that it is not in the simplicity of the ultimate constituents that the answer can be found, but in the richness of the logic by which they move and combine.

Among the many questions that arise:

Why do the particle-points have the same, identical speed?
Why the speed of light?
How do the particle-points combine or aggregate?
How is motion ultimately conveyed or exchanged?
How is it possible for the particle-point to travel at the speed of light irrespective of the frame of reference?
Is acceleration 'absolute' or framework dependent?
What is the nature of the photon and how does it convey motion?
What is the ultimate difference between rest mass and relative mass?

These questions are inter-related. To them we might add what is perhaps the most pertinent question of all: whether or not this scheme is compatible with well established physical theory. Needless to say, the issues are complicated. We will take them up beginning in Part 2 of this essay.

massforceenergy@gmail.com