FUNDAMENTALS

A discussion by Jon Ogborn, emeritus professor of science education at the Institute of Education, London.

The fundamental problem of teaching about energy at secondary level is that school science is obliged to try to run before it can walk. School biology and chemistry need to use the idea of energy before its physical meaning or its measurement in terms of force multiplied buy displacement can be taught.

Teachers want and need to talk about the role energy plays in changes, but the idea that energy is conserved (first law of thermodynamics) is simply not enough to do the job. What they need are some ideas from the second law of thermodynamics.

What *is* energy?

If asked, “Well, what exactly *is* energy, then?”, a lot of people pussyfoot around. You’ll find answers like:

> “energy is the ‘go’ of things”
> “energy is the capacity to do work”
> “energy is a mathematical quantity that is conserved in an isolated system”

The modern answer (the best that can be given so far) is surprising:

> “energy is the source of gravity”

Pick up any nearby massive object, say a paperweight. Heft it in your hand. Feel the downward pull of the Earth on it. Feel how you have to push it to move it to and fro. You are feeling its energy, commonly known as its mass.

This is the meaning of Einstein’s best known equation;

$$E = mc^2$$

A moving object has energy. For Einstein, an object “at rest” is moving, but in time. So as you and the paperweight sit still, you travel towards the future. Your mass, and the mass of the paperweight, is the part of the energy associated with just being at rest, that is, just travelling in time, not in space. So really Einstein’s equation needs to be written:

$$E_{\text{rest}} = mc^2$$

According to Newton’s gravitational theory, mass was the source of the gravitational field. “Masses attract one another”, and all that. According to Einstein, the mass is (normally) just the largest part of the total energy. It is really the energy that gravitates, with the rest energy (alias mass) being by far the largest contribution.

Also, this rest energy can be released, as radiation. If a particle and antiparticle meet, they can annihilate one another, with the rest energy (mass-energy) travelling off carried by a pair of photons.

You see now why other people weren’t telling you what energy really is (assuming they themselves knew). And I haven’t done more than sketch it in here, to give a sense of what the questions are.

Rest energy compared
Because the speed of light $c = 3\ 10^8$ m s^{-1}, the rest energy of a 1 kg mass is:

$9\ 10^{16}$ J

This is about 3 years output of a large 1000 MW power station. Any extra energy the mass has (for example 100 J by being lifted up 10 metre) is a mere flea-bite. An object lifted up against the Earth's gravitational pull does have more energy, and so gravitates more – it is heavier. But by so little that we never notice. The equivalent extra mass is only:

$100\ \text{J}/c^2 = 10^{-15}$ kg approximately.

But it is not nothing. It is roughly the mass of a microbe 1/1000 mm on a side. You could even see the microbe in a microscope. But it is much too small to be detected by simply weighing. You see why energy was not first discovered by noticing the extra weight it could give to objects.

Think of it like this. Everything around us has a huge amount of rest energy (alias mass). This is like a vast ocean of energy. Floating on top, like a thin oil film, is all the extra energy that we notice, calculate, and pay for. Pretty well all the energy that we talk of in science classes belongs to this tiny extra bit.

This 'tiny bit on top' is very important for many practical purposes. But unfortunately this doesn't mean that it has a simple, clear fundamental meaning. The 'real meaning' lies much deeper (as above).

Measuring changes to the 'tiny bit on top' *is* done by measuring amounts of work (force x distance, or some equivalent calculation, such as electric charge x potential difference). This is why energy has often been characterised as 'the capacity to do work'. Nothing wrong with this statement, except that it's rather opaque. Simpler perhaps to say that energy changes are measured in units of work.

Energy is conserved

Energy is conserved. What does this really mean, and why is it true?

Water is more or less conserved. So the amount of water in a reservoir can always be calculated from amount that was there some time ago, *plus* the amount that has come in, *minus* the amount that has gone out (you may have to take account of evaporation as well as water drawn off).

Another way of saying the same thing is that water can't be made or destroyed. For there to be more, it has to come in; for there to be less, it has to go out.

Energy is similar. If you take any volume of space, then the total energy inside that volume at a given time is always the amount that was there earlier, *plus* the total amount that has come in through the surface, *minus* the total amount that has gone out through the surface.

Another way of saying the same thing is that energy can't be made or destroyed. For there to be more, it must have come from somewhere; for there to be less it must have gone somewhere else.

This means that energy can quite correctly be thought of as 'rather like a fluid'. You may correctly picture it as stored or as flowing. You may sensibly ask where it is, where it is going, where it is coming from. [It is *not exactly* like water. For example, you can measure an amount (in J) and a rate of loss or gain (in W). But you can't ask at what velocity the 'stuff' flows, because there isn't any 'stuff' whose 'particles' would have a velocity. We mention this only to keep a few niggles at bay.]

Why is energy conserved? Again, the modern answer is deep and surprising. If the laws governing motion are always the same from day to day – or from aeon to aeon – then there is a corresponding quantity that stays the same. This quantity is the energy. Energy is conserved because 'time of day' doesn't affect any law of motion. This result is fairly easy to state, but much harder to understand. It was discovered (with other similar principles) by the German mathematician Amalie (Emmy) Noether. Her work underlies all modern thinking about conserved quantities.

So there *is* something very abstract about the idea of energy. It is a calculated quantity that must stay constant because there is no natural origin of time. So energy is 'rather like a fluid' *because* it is conserved, not the other way round. It is a calculated quantity, not an observable 'stuff'. The practical teaching implication here is that it is important to do sums about energy changes – how much in, how much out – and not just to talk generally about it.

At the same time, energy is more than just a bit of mathematical machinery. It's real enough for you to be able to feel the attraction between your and the Earth's rest energies.

Energy amongst the molecules

The fundamental fact about chemical and biochemical reactions, in which atoms and molecules re-arrange themselves in different combinations is very simple:

The molecules don't care!

That is, all reactions from rusting through burning to tissue building or muscles contracting occur through purely random behaviour of the molecules, which just happens, on average, to produce the overall desired (or undesired) result. The trick of getting such a reaction to do what you want is to arrange the conditions such that the molecules end up most often doing what you want, just by chance.

Notice that this story (the basis of all thermodynamics) does not yet mention energy. Especially, it doesn't think of energy as "what makes things happen".

Diffusion as an example

When things happen by chance, what happens most often is what can happen in many ways. For example, molecules in gases or liquids diffuse from where the concentration is high to where the concentration is lower. They 'go down the concentration slope'. But they do this simply because where there are more molecules of one kind, there are more of them able to move. And molecules rarely go from where there are fewer to where there are more, because there are fewer around to do so. When the concentrations equalise, molecules go by chance equally in all directions.

Energy flow from hot to cold
The spontaneous thermal flow of energy down a temperature gradient (from hotter to colder) is in many ways like diffusion. In a hot region, molecules have a large average random energy of motion. Colliding with slower moving molecules they are more likely to lose energy than gain it. So the energy concentration tends to equalise, too.

[Energy flow is not *completely* like diffusion. This is because diffusing particles don't change in the process, but 'lumps' of energy aren't 'particles' in the same way. The size of the lumps ('quanta') can change as energy goes from one molecule to another.]

Reactions that release energy
Many reactions or physical changes release energy. This happens when weaker bonds break and stronger bonds form. A physical example is steam condensing or water freezing. A chemical example is combustion. A biological example is the release of energy from adenosine triphosphate in water. [One phosphate ion is detached and clamps hard onto a water molecule, releasing energy.]

There is a tendency for such reactions to go most often in the energy-releasing direction, because the energy released can be shared out amongst neighbouring particles in many ways. To go 'backward' enough energy would need to be collected in one place to pull apart a strong bond, and this will rarely happen by chance if the energy has to be collected from many nearby particles.

Thus energy spreading out amongst many particles is part of what gives a 'driving direction' to exothermic changes.

Explosives are especially dangerous because their reactions also create many smaller particles from fewer larger molecules, as well as releasing energy. The many smaller particles can move in more ways than the fewer larger ones. So this adds to the tendency for the reaction to go in the explosive direction.

Particles getting more organised
It is not true that all energy-releasing (exothermic) reactions happen in the energy-releasing direction. Water freezing is a simple example. When water freezes, stronger bonds form and energy is released. But at the same time, the organisation of the molecules becomes more rigid, less random. So the molecules can be arranged in *fewer* ways. Whether water freezes or not depends on the balance between there being fewer arrangements of molecules and their energy because of the creation of a regular crystal structure, and there being more ways because of the extra energy shared out amongst the particles.

The short way to say this is that the change goes in the direction in which the total entropy goes up. That is, in the direction in which the total number of arrangements of particles and their energy goes up. Depending on the balance, this can be in either direction. For water freezing, the balance changes over at a temperature of 273 K, the freezing point.

Particles getting less organised
Water evaporating is an example of molecules getting less organised – having more space to move about in and more ways to move. That produces a tendency for the change to happen. But at the same time, to free a molecule from the liquid, hydrogen bonds have to be broken and energy has (randomly, just by chance) to arrive in sufficient amount at the right place to do so. This can happen, but not often. It happens less rarely when the water is made hotter, because each particle has more energy of movement. As a result, hot water evaporates, even though this is an endothermic (energy concentrating) change.

As above, these changes go in the direction in which the entropy (number of ways of arranging the 'insides') goes up. In every case, simply because the molecules don't care.

Free energy
'Free energy' is expensive, not free. But it is a useful idea.

Chemists and biochemists often work with the free energy change of a chemical reaction, instead of with the entropy change. The free energy change can be described in two equivalent ways:

1. it is $-T\,\Delta S_{\text{total}}$ where T is the temperature and ΔS_{total} is the total entropy change
2. it is the maximum amount of work available from the reaction

A reaction goes in the direction in which the free energy *decreases* (notice the minus sign, so that this is the direction in which the total entropy *increases*). So free energy is a valuable commodity, which (unlike energy) *is* used up whenever a spontaneous change occurs.

The reason gas, oil and coal are expensive is that they (with the oxygen in the air) provide a source of free energy. The everyday word 'fuel' means pretty much the same as the scientific term 'free energy'. So when we talk about "saving energy" we are talking mainly about saving fuel, that is, free energy.

Similarly, food provides our bodies with a supply of free energy. This makes biochemical reactions, such as tissue building, possible. Thus what is valuable about food is not so much the energy it releases when digested, as the free energy it supplies to drive life-maintaining reactions.

Muscles use free energy deriving from food to contract, so that we can do mechanical work (lifting, running, etc). Reactions in the muscles change myosin molecules so that they bend, dragging an actin fibre sideways. These reactions again involve adenosine triphosphate.

Potential energy is like free energy
The potential energy of a stretched spring or a lifted weight is the work needed to stretch or lift. And all of that work can be got back when the spring relaxes or the weight falls.

If a stone falls on the ground, the energy is shared out amongst many molecules of the ground and stone. The entropy increases. In this kind of case, the potential energy change is the same as the free energy change.

Thus the well known idea that potential energy tends to a minimum is exactly the same idea as that of free energy always decreasing.

Notice that potential energy tends to a minimum only when the energy is dissipated amongst many particles (entropy increases). By contrast, in very simple mechanical examples (a ball on a spring, a planet circling the Sun) there is negligible dissipation. Energy then passes back and forth without loss, from a potential energy store to being stored in the motion (kinetic energy).

Is energy needed for a change to happen?

The usual way round the teaching problem is to talk about the energy "needed for a change to happen". And we get a strong impression from textbooks and everyday talk (not to mention examination specifications and the people who taught us science) that if something has a lot of energy it can make more happen.

These are halfway-useful half-truths.

Certainly if a change involves part of a system increasing in energy, that energy has to come from something else. So evaporating water molecules requires them to be pulled apart, gaining energy, and that energy can come from the thermal store of energy shared out amongst other molecules. Getting a space craft off the ground and into space requires a lot of energy, which comes from burning rocket fuel. If the energy needed isn't there, the change can't happen.

Thus energy conservation (the first law) tells us what can't happen. That's the true part of the half truth.

Lifted weights, stretched springs, fast moving masses, electric currents all have or deliver energy which it may be possible to use to make a change happen. That's because their energy is in effect all free energy, able to drive other reactions by using it up. This is therefore a fairly harmless but not quite right part of the idea of energy as the "go of things".

However energy conservation *can't tell us which way round a change will occur*. For example, the oceans contain a huge amount of thermal energy, but we can't use it to heat our homes or boil kettles. Energy only goes spontaneously from hotter to colder. Thus a hot flame is good for boiling water, because it is *hot*. It is the concentration of energy that matters here.

Similarly, but more subtly, electrical cells and fuels, including foods, all do have the possibility to drive change in a certain direction. They do so because they have a lot of free energy: that is, they have the possibility to increase the entropy enough to balance out another part of a change in which entropy decreases (example – a plant growing tissue out of carbon dioxide and water).

Is there a better way to teach energy?

All this means that teachers do face a big problem. "Energy", which they have to say is conserved, all too obviously does "get lost", because the examples given are really of free energy, which always decreases in any (irreversible) process. So students' commonsense understanding, although in a way exactly correct, runs counter to what they are being taught. In practice, this is resolved by letting the meaning of the term slide backwards and forwards between the two interpretations, which may be good for comfort but not for coherence.

The obvious remedy to get the physics right first, and insist on having both First and Second Law thinking – simply isn't available. To "get the physics right first" would require holding back discussion of energy in biology and chemistry until physics teaching had established the idea of work as the measure of energy transfer, shown that this never changes the total energy, and had discussed thermal transfer of energy as well. That way, you couldn't teach about energy from a biological or chemical perspective much before A-level, and you would probably have put off a lot more students from science. Nobody with any sense advocates such a position.

So, what is to be done? It could help if, as early as possible, pupils played with simple machines like levers to show that although you can magnify a force, or a displacement, you can't magnify both together. More of one means less of the other. It would be good to have less emphasis on word-games with "forms of energy" and more emphasis on where the energy comes from, where it goes, and how it gets there.

It would be very useful if, as early as possible, there were more emphasis on energy calculations. Obvious examples are calculations looking at fuel bills – electricity, oil and gas, and calculations of energy associated with foods. The fact that all can be made to share a common unit can stand for the idea that they are all related to the same thing – amount of energy going from one thing to another, and that energy is in the end a *calculated quantity*. Add in other calculations, such as the potential energy gained by running upstairs, and things begin to fit together.

I'd also like to see more emphasis on dissipation, understood as the spreading out of energy amongst vast numbers of atoms or molecules. It is needed to make sense of the conservation of energy, and to understand the direction of processes.

School science should also have some simple version of Second Law thinking to hand. Ultimately what matters is *differences or gradients*; differences of temperature, of concentration of matter, of potential energy. The idea that (irreversible) processes always "go downhill" seems to me to be viable as a piece of teaching. It also offers a language that could be in common between all the sciences, making it much less important who does what first.

