Entropy

Anindya Chatterjee
Mechanical Engineering, IIT Kanpur

November 17, 2025

Entropy is difficult to understand. I think it is an amazing achievement of classical thermodynamics: the idea that such a quantity *must necessarily* be a state variable. Last year, through discussions with Uma Chatterjee, my understanding became clearer. So I am writing these notes. I have aimed for brevity and clarity.

Let us first get the "entropy is disorder" distraction out of the way. There is classical thermodynamics, where the existence and necessity of entropy was deduced independently (and in my opinion brilliantly). And there is statistical thermodynamics, where (also brilliantly) something was defined in a different way and observed to be the same as the classical entropy. In the statistics of ensembles there is sophisticated reasoning about the system preferring to be in configurations that are much more likely than other configurations, and there is clever mathematics behind the "much more likely" part. Classical thermodynamics discovers entropy independently and differently. I think that people studying classical thermodynamics should understand the classical definition of entropy without taking the tempting shortcut of saying things like, "Oh, it is just disorder, and obviously disorder tends to increase." In this latter reasoning the distinction between precise science and loose social constructs can be lost if one is careless. The connection between this disorder and heat and temperature and the second law are, in my opinion, not understood by most people who say entropy is disorder.

So, classical thermodynamics, then.

A thing to note is that even classical thermodynamics is new compared to mechanics. From mechanics we know (or at least know how to think about) things like force, mass (inertia), distance, and work. These tie in well with our lived experience.

For thermal phenomena, our lived experience lets us know if one thing is hotter than another thing. *That is all.* From this experience we have an idea of how two bodies can be in contact with neither of them being hotter than the other: they are in thermal equilibrium. The property they then share is called temperature. Note that we have not yet defined what temperature is. We merely know when two bodies are at the same temperature and when one body is hotter than the other. It is a side issue that someone made a mercury in glass thermometer, found the length of mercury in the tube changing with temperature, defined two points on the scale (melting ice and boiling water), and used a uniform scale between these points. That is empirical, dependent on the properties of mercury and glass, does not guarantee extension outside that range, and does not imply an absolute zero in the temperature scale. For these reasons, my position at this point is that we know there is a thing called temperature but we have not yet tied ourselves down to a scale for measuring it.

We do not understand heat yet. We cannot catch heat and measure it. Its effect on a body is to change its temperature. And so, it is a big moment when we do work on a system, like turn a paddle inside a tank of water that is externally insulated, and find that the water gets hotter when we turn the paddle (note that we understand insulation from lived experience because we have been wearing warm clothes for a long time). We conclude that work can be converted to heat, and assign the *same units* to them. In principle we could quantify an amount of heat through the work needed to generate it.

Let me clarify the last line. Although we do not have a temperature scale yet, let two identical bodies start in thermal equilibrium. We insulate them from the universe and from each other. Now one body is

deliberately given some amount of heat Q which we wish to measure. The body becomes hotter. We then do work on the other body until it reaches the same temperature as the first body. We measure the work we have done (we know how), and now we have indirectly measured Q. In this way, in principle, we have a scale for heat and we know how to measure it too.

We introduce a construct called internal energy. When we heat a body, its internal energy increases. We may have ideas in the background that involve vibrating molecules, but that is not a part of the classical subject. We acknowledge that things like batteries have internal energy stored in chemical form, and so internal energy can be more than just stored heat. If we have machines that involve lifting weights as well as heat exchange, then internal energy could include gravitational potential energy. We trust that, from system to system, we will understand what is meant by the internal energy of the system. I think most people have no trouble in "understanding" internal energy the way they have trouble in "understanding" entropy. So, I say we now know what internal energy is, and it has the same units as work and heat. Note that internal energy is a state variable: changes in it depend on the change in state and not the path followed. Contrast with a box we slide around on a frictional floor along a complicated path, restoring the box eventually to its original position. The work done depends on the path followed.

Now we think of a process.

A minor statement on a major matter: reversibility. If we do things slowly enough, many processes are reversible at least in principle. Compression of gases, pumping of liquids to heights, heating of bodies, and so on. For simplicity, I have not dwelled below on the reversibility aspect. Wherever a process is described below that can be reversible in principle, assume that it is reversible.

In a process, suppose we supply heat ΔQ to a system. It produces work output ΔW and its internal energy increases by ΔU. Failure to build perpetual motion machines has taught us the first law:

$$\Delta Q = \Delta W + \Delta U.$$

In particular, if the process is a cycle, i.e., the system state at the end is the same as it was at the start, then

$$\Delta U = 0$$

and

$$\Delta Q = \Delta W \text{ in a cycle.} \tag{1}$$

That equation above explains why a lot of would-be perpetual motion machines will not work. You cannot steadily (in a cycle) get work output that exceeds energy input.

We still do not have a scale for temperature. But we know when something is insulated (no heat exchange) and when a process occurs at constant temperature. The first type of process is called adiabatic, and the second type is called isothermal.

We now come to a second type of would-be perpetual motion machine. This type of machine absorbs heat and turns all of it into work, obeying Eq. 1. A machine that turns supplied heat into work is called a heat engine. Failure to build perpetual motion machines has taught us that, loosely speaking, a heat engine working in a cycle cannot convert all heat supplied into work. This sentence is loose because Eq. 1 is fundamental. What I mean is that we distinguish between *supplied* heat and *rejected* heat. The *net* heat added, ΔQ, is the supplied heat minus the rejected heat, and Eq. 1 applies to ΔQ. The second law says that the net heat added must have a supplied heat part but also a nonzero rejected heat part. One statement of the second law is this (Clausius statement):

> It is impossible to construct a device whose sole effect is the transfer of heat from a cold body to a hot body.

An implication is that heat will not flow spontaneously from a cold body to a hot body. And that, to transfer heat from a cold body to a hot body, nonzero work must be done. There is another statement of the second law (Kelvin-Planck statement):

> It is impossible for a heat engine working in a cycle to produce nonzero net work output while exchanging heat with only one reservoir[1].

This means there must be at least two reservoirs at two different temperatures. The reservoir at a high temperature supplies heat to the heat engine. The heat engine converts some of the supplied heat to work, but it must necessarily reject some nonzero remaining portion of the heat to the lower temperature reservoir.

The Clausius statement of the second law implies that a reversible engine operating between two reservoirs at different temperatures must produce nonzero work output (because, otherwise, we could run it in reverse and violate the law). We will use this below.

As an aside, for clarity, it may help to think of water and steam in a thermal power plant. You start with water at a somewhat low pressure. You pump the water to a high pressure and send it to a boiler. A fire in the boiler heats the water and turns it into superheated steam. The steam goes through a turbine which does external work (it drives a generator which feeds electricity into the grid). The steam exiting the turbine is cooler and at a lower pressure. You must now condense it back to water, rejecting the latent heat (to a river or lake, perhaps), before feeding the water back to the pump. Defining ΔQ to be the heat supplied in the boiler minus the heat rejected to the river; and defining ΔW to be the turbine work output minus the pump work input; Eq. 1 applies. The second law does not negate the first law; it merely constrains the possibilities further.

The second law has the following remarkable consequence. Consider two heat engines working in cycles between the same two reservoirs. One engine, E_1, is reversible; and the other, E_2, may or may not be reversible. Then the efficiency of E_1 is not less than the efficiency of E_2.

To see why, let us argue by contradiction. Suppose the efficiency of E_1 is indeed less than that of E_2. Instead of symbols for generality, we can use numbers for a specific example (the reader can restate the argument using symbols). Imagine that the efficiency of E_2 is 0.5, and the efficiency of E_1 is 0.4. Let 100 J (Joules) of heat be supplied to E_2 from the high temperature reservoir. Then E_2 provides 50 J of work output and rejects 50 J to the lower temperature reservoir (E_2's efficiency is 0.5). A simple calculation shows that if 83.33 J of heat is supplied to E_1 by the high temperature reservoir, then E_1 provides $0.4 \times 83.33 = 33.33$ J of work output and rejects 50 J of heat to the lower temperature reservoir. Now we can run E_1 in reverse (E_1 is reversible). Upon reversal, E_1 absorbs 50 J of heat (the same as the heat rejected by E_2), requires 33.33 J of work input, and pumps 83.33 J of heat into the high temperature reservoir. The heat rejected by E_2 can directly go to E_1, eliminating the low temperature reservoir. The combination of E_1 and E_2, together considered as *one* heat engine, produces $50 - 33.33 = 16.67$ J of net work output, and absorbs $100 - 83.33 = 16.67$ J of heat from the high temperature reservoir without rejecting any heat to a lower temperature reservoir (i.e., it exchanges heat with only one reservoir). We have violated the second law, and obtained a contradiction. Thus, our initial supposition must be false. The efficiency of E_1 is not less than that of E_2. The line of argument above applies even if E_2 is reversible. It follows that no reversible engine working between two given temperatures can have an efficiency lower than that of any other reversible engine working between the same two temperatures. All reversible engines operating between the same two temperatures have the same efficiency.

We still have not defined a temperature scale. We can now do that. In working through the reasoning of the following paragraph, the reader might use a pen and paper to make sketches for consolidating ideas.

Let a reversible heat engine E_1 operate between two reservoirs at temperatures T_1 and T_2, with $T_1 > T_2$. The heat engine is supplied heat Q_1 at T_1, produces work output W_1, and rejects Q_2 at T_2. Its efficiency is

$$\eta = \frac{\text{work output}}{\text{heat supplied}} = \frac{Q_1 - Q_2}{Q_1} = 1 - \frac{Q_2}{Q_1}.$$

If we hold T_1 and T_2 constant, then Q_2/Q_1 is fixed regardless of the amount of heat Q_1 as well as the detailed construction of the engine. Now consider a still lower temperature, $T_3 < T_2$, and a second reversible heat

[1] A reservoir can absorb or supply heat without any change in its own temperature. Think of a big lake if you like. From the reservoir's imperturbable point of view, heat transfer to and from it is always reversible.

engine E_2 that operates between T_2 and T_3. E_2 also produces nonzero work output. Suppose E_2 is supplied heat Q_2 (the same as rejected by the first engine, at the same temperature T_2), produces work output W_2, and rejects heat Q_3. Then $Q_3 < Q_2$, because $Q_3 = Q_2 - W_2$, and so

$$\frac{Q_3}{Q_1} < \frac{Q_2}{Q_1}.$$

We can consider the two reversible heat engines E_1 and E_2 to be a combined effective reversible heat engine operating between T_1 and T_3. In this way, holding T_1 and Q_1 fixed arbitrarily at any value we like, we find that if we lower T_2 then Q_2 decreases. Q_2 and T_2 are monotonically increasing functions of each other. Since the actual temperature scale is so far arbitrary, we can now *define* the temperature scale to be such that

$$\frac{Q_2}{Q_1} = \frac{T_2}{T_1}.$$

The temperature scale has now been defined upto an arbitrary multiplicative constant. Such a scale can be called an *absolute* temperature scale because it is defined using fundamental laws (the first and second laws of thermodynamics) and does not depend on specific materials (ice, water, glass, mercury) and specific instruments (bulb and tube thermometers).

The arbitrary multiplicative constant of the previous paragraph deserves some discussion. Suppose we start at some high temperature T_1, like (say) the melting point of sodium. We build a reversible heat engine (possible in principle). We run it in a cycle such that the heat supplied, Q_1, is some known quantity, like (say) 1 kJ. We use melting ice (which we empirically think of as 0°C) as our lower temperature reservoir. We measure Q_2, the heat rejected (possible in principle). Now we can choose an arbitrary positive number, like 100 or 540 or 273, and let the lower temperature T_2 in the absolute scale be that number. For each such choice, we have a corresponding temperature scale. Of these choices, we find the number 273 convenient because then increments of 1 in our new absolute temperature scale match increments of 1°C in our familiar empirical temperature scale. As far as I know, this match is a nice coincidence and not predictable from the theory developed so far. The new absolute temperature scale thus obtained is called the Kelvin scale: 0°C is the same as 273 K. (A more precise value is 273.15. We will not get into how it is measured.)

We have a name for a reversible engine that operates between two temperature reservoirs. We call such an engine a Carnot engine. The efficiency of a Carnot engine is, by definition of the temperature scale,

$$\eta_{\text{Carnot}} = 1 - \frac{Q_2}{Q_1} = 1 - \frac{T_2}{T_1}. \tag{2}$$

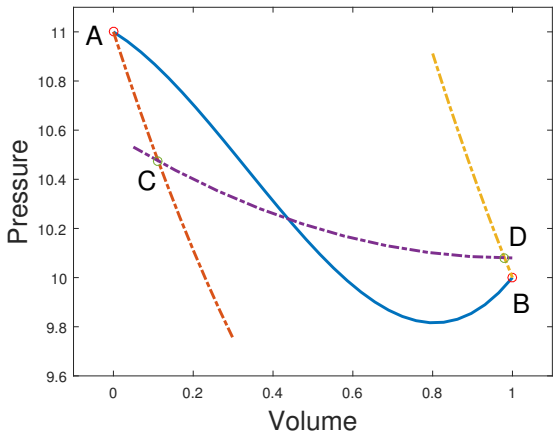
It seems clear now that a Carnot engine must work in a cycle with 4 fundamental processes. From state 1 to state 2 we have constant temperature or isothermal heat addition Q_1 at T_1. From state 2 to state 3 we have a work output process, which is adiabatic. From state 3 to state 4 we have isothermal heat rejection Q_2 at T_2. From state 4 to state 1 we have adiabatic work input, completing the cycle. From Eq. 2 we have

$$\frac{Q_1}{T_1} = \frac{Q_2}{T_2}.$$

However, if we think of dQ as heat supplied, and assign a negative sign to heat rejected, then for such an engine we have for the cycle

$$\oint \frac{dQ}{T} = 0. \tag{3}$$

We now come to an interesting result. See the figure below (ignore the numbers on the axes, they are arbitrary).



Taking pressure and volume of the working medium in a heat engine as state variables, consider an arbitrary reversible process from state A to state B, as shown by the solid line.

Now adiabatic processes on these axes are steeper than isothermal processes. (One way to see it is to think of a gas. You let it expand without heat exchange, and its pressure and its temperature both decrease. That is an adiabatic process. Now you hold the lowered pressure constant and heat it back to its initial temperature, and the volume increases. The net curve for the isothermal process is, thus, less steep.)

With the above, in the figure, consider moving from state A to state B in three steps. The first step is an adiabatic process from A to C. The second step is an isothermal process from C to D. The third step is an adiabatic process from D to B. The specific isothermal curve chosen is the one that makes the net work done in process ACDB equal to the net work done in the original process AB. Now since endpoints A and B are the same for both processes, the change in internal energy is the same. The work done is the same by construction. Thus, by the first law, the heat supplied is the same as well:

$$\Delta Q_{AB} = \Delta Q_{ACDB} = \Delta Q_{CD}, \tag{4}$$

because the adiabatic parts of the process do not involve heat exchange. Note that ΔQ_{CD} is added at a constant temperature. In this way the process AB has been replaced with an equivalent three-step process ACBD, wherein the change in internal energy is the same, the start and end points are the same, the heat added is the same, and the work done is the same. Note that the change from A to B could be small, even infinitesimal. If there is a general reversible process where the net change is large, we can split it into a sequence of small ones. Then each small one can be treated like process AB above.

We now need another figure. A figure from somebody's lecture notes[2], adapted to show one small cycle more clearly (in red), is given on the next page.

The figure shows how an arbitrary reversible cycle can be approximated by a large number of small Carnot cycles. The jaggedness of the approximating boundary does not matter if the width of each cycle goes to zero because, by construction, all changes in relevant quantities are the same. For each small cycle, we can identify an upper and a lower boundary, both of which are isothermal (see the relevance to the previous figure and to Eq. 4). For each small cycle, we have Eq. 3. Therefore, for the entire arbitrary reversible cycle, we have Eq. 3 as well, which is worth writing down again:

$$\oint \frac{dQ}{T} = 0.$$

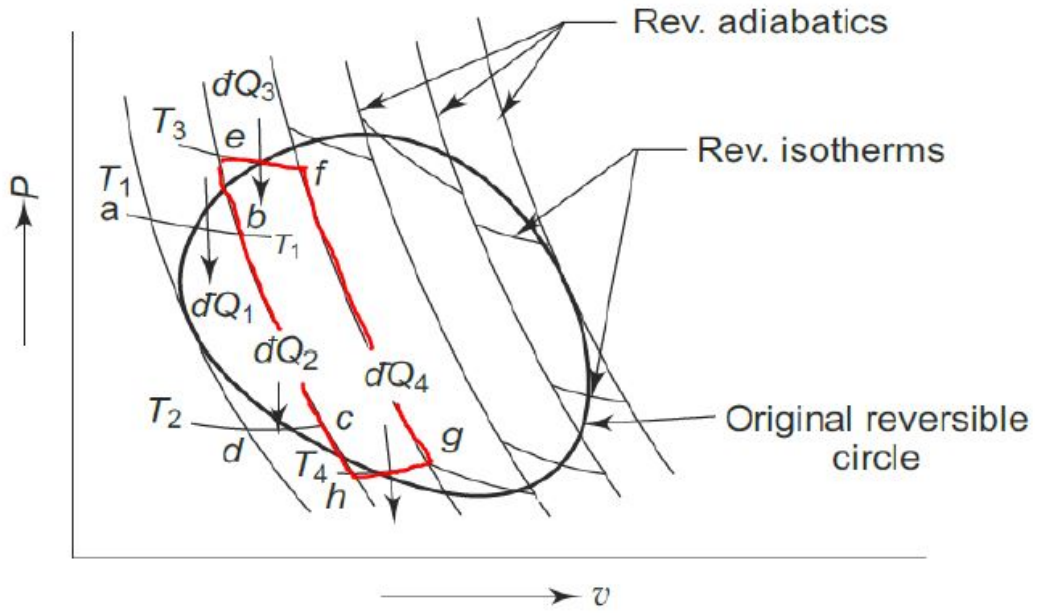
If we now define a quantity S and call it entropy; and we say that the change in entropy in any reversible process of heat addition is

$$dS = \frac{dQ}{T},$$

[2]The figure is apparently copied there from somewhere else: I do not know the original source. To avoid embarrassing the person who made the slides I am not naming them here.

then in any reversible cycle

$$\oint dS = 0.$$



This means entropy must be a state variable. In classical thermodynamics we cannot find its actual or absolute value, which may remind us of potential energy. But, if we choose a datum or reference value, we can compute the changes in it: again, just like potential energy. Note that entropy is not energy: it has units of Joules per Kelvin, J/K.

Note that if a hot body at temperature T_1 is in contact with a cooler body at temperature T_2 and an infinitesimal quantity of heat dQ flows irreversibly from the hot body to the cold body, then the entropy of the hot body changes by $-\dfrac{dQ}{T_1}$ while that of the cooler body changes by $\dfrac{dQ}{T_2}$. Considering both bodies together, the net entropy change in this irreversible process is

$$\frac{dQ}{T_2} - \frac{dQ}{T_1} > 0$$

because $T_1 > T_2$. In this way, irreversible processes are accompanied by an increase in entropy. Positive entropy changes drive spontaneous processes.

We can close with a simple example that illustrates why, for a system undergoing an irreversible cycle,

$$\oint \frac{dQ}{T} < 0.$$

The above may seem initially surprising because we have heard that "entropy always increases". So, consider a heat engine E, a hot reservoir R_1 at temperature T_1, a low temperature reservoir R_2 at temperature $T_2 < T_1$, and an external machine M. R_1 supplies heat Q_1 to E, E produces work output W that is absorbed by M, and E rejects heat Q_2 to R_2. The *system* is E, and its *environment* comprises R_1, M and R_2. Now, first suppose that E is reversible and works in a cycle. For E,

$$\oint dS = \oint \frac{dQ}{T} = \frac{Q_1}{T_1} - \frac{Q_2}{T_2} = 0$$

as discussed in detail in these notes. Next, suppose Q_1 is held fixed, but E is not reversible. Then Q_2 is greater than before. As a result, for E, we still have

$$\oint dS = 0$$

because it works in a cycle and entropy is a state variable; but for E now,

$$\oint \frac{dQ}{T} = \frac{Q_1}{T_1} - \frac{Q_2}{T_2} < 0$$

because Q_2 has increased. The reader may like to look at the above two integrals again. For an engine working in an irreversible cycle, the first integral remains zero but the second does not. For the environment it is not meaningful to ask if

$$\oint dS = 0$$

because the environment does not operate in a cycle: it loses heat and receives work. In fact, for the environment,

$$\int dS = \frac{Q_2}{T_2} - \frac{Q_1}{T_1} > 0.$$

For the universe, i.e., for the system plus its environment, over one cycle of the irreversible engine E, the net change in entropy is positive.

We now understand within the classical framework why entropy *must* be a state variable, *what* it must be, and (at least in principle) *how* it must change for reversible and irreversible process. I stop here.

Acknowledgement: Sankalp Tiwari provided useful editorial inputs. Surviving errors remain mine.