Lecture delivered at Thomas Aquinas College

The Discovery of Entropy and its Significance

Adam Schulman
St. John's College, Annapolis

November 3, 2023

My theme is entropy—how it was discovered, what it means, and what might be its wider and deeper implications. I think every educated human being ought to know something about entropy and the Second Law of Thermodynamics, which says that the entropy of the universe, or of any isolated part of it, is always increasing. After all, aside from Darwin's idea of natural selection, the relentless increase of entropy is perhaps *the* discovery of modern science that has most profoundly colored our current view of human life and its status in nature as a whole. Parenthetically, it seems to me that most people find the *First* Law of Thermodynamics, which says that energy is always conserved, to be a more or less clear and reasonable principle, with little of the mystery and obscurity surrounding entropy and the Second Law. That attitude may not be justified, and at the end of this talk I will suggest that the opposite might be true.

I. Energy and Entropy, Optimism and Pessimism

So, this is a talk about entropy; but in a way it is also about optimism and pessimism in modern thought, as I will show you shortly.

The modern scientific notion of *energy*, or the capacity to do work, was first discovered and championed by Leibniz, the German mathematician and philosopher, at the end of the seventeenth century. He used the Latin expression *vis viva* or living force, but since about 1850 we have called it "energy," derived from the Greek *energeia,* a word coined by Aristotle to mean something like activity. Leibniz was the first to propose the conservation of energy as a fundamental law of nature.

Around the same time Leibniz also expressed his view of the ultimate fate of the universe. In 1697, in an essay "On the radical origin of things," he wrote:

> …[W]e must recognize a certain perpetual and very free progress of the universe as a whole, so that it always proceeds toward greater culture.… And though it may be objected that, if this were so, the world should have become a paradise long ago, there is an answer near at hand. Although many substances have already reached great perfection, yet…there always remain in the abyss of things parts that are still asleep, yet to be aroused and advanced into something greater and better and, in a word, to a better culture. Thus progress never comes to an end.

Such was the optimism that could accompany the discovery of energy conservation at the end of the seventeenth century.

Now what Leibniz discovered was, more precisely, the interconvertibility of *kinetic energy*, or *vis viva*, the kind of energy that moving bodies possess, and *potential energy*, the kind that is stored up when a spring is compressed or a weight is elevated above the surface of the earth. It took another 50 years before modern science recognized that *heat* is yet a third form of energy, interconvertible under certain circumstances with kinetic and potential energy. The leading figure in that discovery was the English brewer-turned-physicist James Joule, who like Leibniz was a confirmed optimist. Here he is commenting on the cosmic implications of the inclusion of heat in the conservation of energy: "Behold, then, the wonderful arrangements of creation…." Despite the "apparent destruction of living force" in "almost all natural phenomena," "we find that no waste or loss of living force has actually occurred." "Thus it is that order is maintained in the universe—nothing is deranged, nothing ever lost, but the entire machinery, complicated as it is, works smoothly and harmoniously." That is Joule in 1847, perhaps the high water mark of thermodynamic optimism.

On April 24, 1865, ten days after the assassination of President Lincoln, a new word came into being. That word is *entropy*. It was the invention of another German scientist named Rudolph Clausius. He fashioned it from the ancient Greek word *trope*, meaning "transformation." As Clausius explained, he wanted to form a word that would be as similar as possible to the word "energy;" and he borrowed its root from the ancient Greek in the hope that the word would be adopted unchanged in all modern languages. That hope has been abundantly realized.

Like energy, entropy is the subject of a fundamental law of nature. But while energy can neither be created nor destroyed, entropy can only increase. As Clausius put it, summing up the two laws, "The energy of the universe is constant, the entropy strives toward a maximum."

Energy and entropy are the two basic notions of thermodynamics, the science of heat and its transformation into mechanical work. Energy, as I have indicated, may be defined as the capacity to do work. Because of the conservation of energy, it is not possible to do work without expending energy. This rules out the possibility of a "perpetual motion machine" i.e., a machine that produces more energy than it consumes and thus can run indefinitely without further consumption of energy. Some would say that the *intuition*—or if you like the *axiom*—that no such machine is possible is the logical ground for the First Law of Thermodynamics.

Heat, as we have known since Joule, is also a form of energy, namely, thermal energy. But thermal and mechanical energy differ in an important respect. It is always possible to transform mechanical energy into thermal energy, but it is not always possible to transform thermal energy into mechanical energy. For example, a spinning propeller immersed in water will heat the water up by friction. But the heat in the water cannot in general be transformed back into the motion of a propeller. In fact, most heat is not available to do mechanical work at all. The Atlantic Ocean, for example, is a vast reservoir of heat, but very little of that heat can be used to do work.

The entropy of a system is, roughly speaking, an indication of how much heat is available to do mechanical work. If the entropy increases, that means there is less available heat. And if, as Clausius stated, the entropy can never decrease, it follows that any heat that has by an increase of entropy become unavailable for work is irrecoverable. Applied to the universe as a whole, the law of the increase of entropy has a rather disturbing consequence, which was first noticed in 1852 by the British physicist William Thomson, later Lord Kelvin. Kelvin was the first to spell out the cosmic implications of the emerging science of thermodynamics. Here, in part, is the ominous conclusion of Kelvin's paper:

> There is at present in the material world a universal tendency to the dissipation of mechanical energy…Within a finite period of time past, the earth must have been, and within a finite period of time to come the earth must again be, unfit for the habitation of man as at present constituted….

Thus we see, within a few generations, the sunny optimism of Leibniz and the age of reason giving way to the doom and gloom of the nineteenth century, when scientists could prophesy with confidence the ultimate heat death of the universe. It seems to me that this cosmic pessimism, which began to take root in the late 19th century, has since become the dominant mood of the West. I am not of course insisting that the discovery of entropy bears primary responsibility for this development, but it must have helped. Here, for example, is Charles Darwin in 1876, commenting on "the view now held by most physicists, namely, that the sun with all the planets will in time grow too cold for life, unless indeed some great body dashes into the sun, and thus gives it fresh life." Darwin continues: "Believing as I do that man in the distant future will be a far more perfect creature than he now is, it is an intolerable thought that he and all other sentient beings are doomed to complete annihilation after such long continued slow progress."[1]

Next, here is the British logician and gadfly Bertrand Russell, writing in 1903:

> That Man is the product of causes which had no prevision of the end they were achieving; that his origin, his growth, his hopes and fears, his loves and his beliefs, are but the outcome of accidental collocations of atoms; that no fire, no heroism, no intensity of thought and feeling, can preserve an individual life beyond the grave; that all the labours of the ages, all the devotion, all the inspiration, all the noonday brightness of human genius, are destined to extinction in the vast death of the solar system, and that the whole temple of Man's achievement must inevitably be buried beneath the debris of a universe in ruins—all these things, if not quite beyond dispute, are yet so nearly certain, that no philosophy which rejects them can hope to stand. Only within the scaffolding of these truths, only on the firm foundation of unyielding despair, can the soul's habitation henceforth be safely built.[2]

And finally, closer to home, here is the American theoretical physicist Steven Weinberg, in the closing paragraphs of his 1977 book, *The First Three Minutes*:

> It is almost irresistible for humans to believe that we have some special relation to the universe, that human life is not just a more-or-less farcical outcome of a chain of accidents reaching back to the first three minutes, but that we were somehow built in from the beginning…. It is very hard to realize that [our world] is just a tiny part of an overwhelmingly hostile universe. It is even harder to realize that this present universe has evolved from an unspeakably unfamiliar early condition, and faces a future extinction of endless cold or intolerable heat. The more the universe seems comprehensible, the more it also seems pointless.

---

[1] Charles Darwin, "Recollections of the Development of my Mind and Character" [Autobiography], written August 3, 1876, in *The Life and Letters of Charles Darwin,* ed. Francis Darwin (London: John Murray, 1887).

[2] Bertrand Russell, *The Free Man's Worship*, December, 1903.

Weinberg adds some words of comfort:

> But if there is no solace in the fruits of our research, there is at least some consolation in the research itself…. The effort to understand the universe is one of the very few things that lifts human life a little above the level of farce, and gives it some of the grace of tragedy.

It is partly in this spirit that I offer the following account of the discovery of entropy. By the end of this hour I hope to have exhibited the difference between merely knowing by hearsay that the entropy of the universe is always increasing, and understanding with some precision where that idea came from, what it really means, and what evidence it rests on. The remarkable story of its discovery spans the years 1824 to 1854, during which most of classical thermodynamics was created by three men: one French, one British, and one German.

II. Carnot

Our story begins in 1824, when a young French engineer named Sadi Carnot wrote a long essay entitled *Reflections on the Motive Power of Fire*. With this little book was born the science of thermodynamics. That science reached maturity essentially within one generation, thanks largely to the work of Kelvin and Clausius.

Thermodynamics is the theory of the relation between heat and work, but its origins lie in practical engineering. The early nineteenth century was the age of the steam engine. In a steam engine, combustible fuel is burned in a furnace, water is thereby boiled, and the pressure of the expanding steam is used to drive a piston. A steam engine is an example of a heat engine, which expends a certain quantity of heat to produce a certain quantity of usable work. Steam is known as the expansive agent or working substance in a steam engine; other heat engines use other substances, such as air, alcohol, etc. Carnot's *Reflections* begin with a series of questions chiefly of interest to engineers: What is the most efficient means of deriving work from heat? Is there any limit to the motive power of fire, i.e., of the ability of heat to do work? Could improvements to the steam engine raise without limit the ratio of work produced to fuel expended? Would a heat engine be more efficient, for example, if it worked with an expansive agent other than water vapor, such as air? Or is there an assignable limit to the motive power of fire, "a limit which the nature of things will not allow to be passed by any means whatever"? On these seemingly modest practical foundations Carnot built the theoretical edifice of classical thermodynamics.

At first glance, it might seem that a hot body all by itself can always be used to do work. One simply puts the hot body in contact with an expandable vessel filled with a gas such as air or steam and uses the pressure of the expanding gas to drive a piston. Voila, work has been done. But Carnot was the first to see clearly that to derive work from heat one must have not only a hot body but also a cold body. To understand his insight, it will be helpful first to say something about the context in which Carnot was thinking and writing about heat engines.

At the end of the eighteenth century the chief rival to steam power was water power. Consider how a water engine—that is to say, a water wheel—derives work from the water flowing down a stream. A water wheel generates motive power by taking in water at a given height and releasing it at a lower height. Work can be derived from water engines only where there is water elevated above the adjacent terrain. As we would say, the potential energy of the elevated water is converted into the mechanical energy of the water wheel. Or, as Carnot and his contemporaries would put it, the fall of water is the source of the motive power of the water engine.

If all the water on earth were at sea level, there would be in effect a kind of equilibrium, and no work could be done by water engines. Only where there is a *lack of equilibrium* in the elevation of water, i.e., wherever water is raised above some of its surroundings, can work be done when the water falls.

In reflecting on the steam engine Carnot noticed what seemed to him an analogous fall of heat. What happens in a steam engine is this: heat is generated at the furnace at a high temperature; it is then incorporated into the steam which expands and drives the piston; then the steam is brought in contact with a cold body or refrigerator, which condenses the steam by cooling it, after which it is again put in contact with the furnace, and the cycle begins all over again.

Carnot saw that, in every case, work can be produced from heat only if some heat is absorbed from the furnace at a higher temperature and discharged at the refrigerator at a lower temperature. He concluded that, just as the fall of water from one height to another is the source of motive power in the water wheel, so too the fall of heat from one temperature to another is the source of power in a steam engine, and hence presumably in any heat engine.

Now it may have occurred to some of you that one could operate a steam engine without any refrigerator at all. That is, after boiling the water and using the pressure of the expanding steam to drive the piston, you simply open the piston, expel the steam into the atmosphere, add new water, and begin the boiling process all over again. In fact, the more primitive steam engines of Carnot's time did operate in just this way, using fresh water for each cycle and eliminating the condensation phase. In that case, what becomes of Carnot's fall of heat?

Carnot, however, anticipated this objection and answered it by pointing out that in such instances the atmosphere itself is cooler than the steam and functions as an enormous refrigerator. That is, when the piston is opened, the heated steam goes away only because the surrounding air is cool enough to receive it. Thus the condensation phase cannot really be eliminated. In harnessing the motive power of heat, one cannot escape the necessary flow of heat from a hot body to a cold body.

As Carnot put it, the motive power of a steam engine is due to a "re-establishment of the equilibrium of caloric." That is, just as water spontaneously flows downhill to restore its mechanical equilibrium, so heat spontaneously flows from a hotter body to a colder body to restore its thermal equilibrium. Just as water engines could do no work if all water were at sea level, no work could be done by steam engines if all bodies were at the same temperature. And just as water never spontaneously flows uphill, heat never spontaneously moves from a colder body to a hotter body. Conversely, *wherever there exists a difference of temperature, motive power can be produced,* by allowing the heat to fall from the hot body to the cold body while driving a heat engine.

Carnot took the analogy with water power one step further, however. Just as the water driving a water wheel is not destroyed in passing from one elevation to another, Carnot reasoned that heat is not consumed in passing through a heat engine; it is merely transported from a hot body to a cold body. Carnot found it easier to draw this conclusion because he believed in the conservation of heat. That is, he subscribed to the common view, championed by Lavoisier in the previous century, that heat is a subtle material fluid, usually called caloric, that can be neither created nor destroyed. The very title of Carnot's book, *On the Motive Power of Fire*, already indicates his belief in the material nature of heat. The fall of heat that produces motive power in a steam engine is thus the flow of caloric from a hot body to a cold body. And since heat or caloric is indestructible, the steam engine does not operate by converting heat into work; in the cyclical operation of the steam engine, the same quantity of heat that is absorbed by the steam at the furnace must also be discharged by the steam at the refrigerator.

By the 1850s it was to become clear that the caloric theory was untenable as an account of the nature of heat, and it began to give way to the rival mechanical theory of heat. According to the mechanical theory, heat is *not* a subtle material substance, it is a form of motion. Heat is *not* conserved during the operation of a steam engine, it is consumed in the production of work; indeed, the quantity of work produced by the consumption of a unit of heat is always the same, a universal constant of nature. Thus the steam engine produces a certain quantity of work only because a definite quantity of heat is consumed during each cycle. Accordingly, on any cycle in which work is done, the quantity of heat discharged at the refrigerator must be smaller than the quantity of heat absorbed at the furnace. That is, of the heat absorbed at the furnace, part is converted into work, and part is passed to the refrigerator.

Note that the old caloric theory and the new mechanical theory of heat are not in disagreement about the conservation of energy. Carnot himself accepted the principle that energy, the capacity to do work, can be neither created nor destroyed. What Carnot denied is that heat is a form of energy; instead, he considered it an indestructible substance that is capable of doing work when it undergoes a fall in temperature, just as water is capable of doing work when it undergoes a fall of elevation.

Overall, Carnot had three brilliant insights that survive to this day, unsullied by his error regarding the nature of heat. The first insight is the fall of heat, which we have already considered. The second insight is the cycle: Carnot saw that, in order to understand the motive power of a steam engine, it is not sufficient to look only at the heat generated in the furnace and the work done by the piston. In order to see the relation between heat and work one must contrive to return the heated and expanded steam to its original condition. That is, the steam must be cooled and compressed so that it is in exactly the same state it was in before it absorbed heat from the furnace. Only if the cycle is completed in this way can we be confident that the work done by the piston is the only mechanical effect of the heat absorbed at the furnace. If we do not complete the cycle, but leave the steam in its heated and expanded state, we would have to include the alteration of the state of the steam as an effect of the heat absorbed at the furnace. In short, Carnot saw that a complete cycle must be considered if we are to understand fully the necessary and sufficient conditions for heat to do work.

Consideration of the complete cycle and the fall of heat led Carnot to his third great insight. He noticed that under certain circumstances the operations that make up the cycle are reversible. What do we mean by a reversible process?

Consider again the water wheel. Under normal operation the wheel produces work by taking in water at the upper level and discharging it at the lower level. But the wheel could be operated in reverse, taking in water at the lower level and discharging it at the upper level. In that case, of course, we would have to supply the work of turning the wheel and raising the water. In effect, a water wheel operating in reverse is a machine for converting mechanical energy into the potential energy of water at a raised elevation in the earth's gravitational field.

Note, however, that no actual water wheel can be entirely reversible. If, at any point in the operation of a water wheel, any water is allowed to fall some distance without driving the wheel, that part of the operation is irreversible, because turning the wheel in reverse would not restore the water to the height that was lost. In a perfect or completely reversible water wheel, the water must not be permitted to fall any distance without driving the wheel as it falls. Of course, water wheels fall short of perfect reversibility in other ways as well: any friction in the operation of the machinery, any turbulence in the flow of the water, are processes that it would be impossible to reverse if the wheel is operated backwards.

Carnot noted that heat engines, too, can be operated in reverse. Under normal operation, a heat engine produces work by taking in heat at the furnace at a higher temperature and discharging heat at the refrigerator at a lower temperature. Operated in reverse, a heat engine would absorb heat at the lower temperature and discharge heat at the higher temperature. In effect, a heat engine operating in reverse is a refrigerating machine (or a heat pump), which extracts heat from a cold body and expels it into a hotter body. Of course, to operate a refrigerating machine we would have to supply work in order to drive the heat engine in reverse. In a reverse steam engine, for example, we would have to compress the steam mechanically at the higher temperature so as to expel its heat to the hotter body or furnace.

Carnot saw that not every heat engine can be operated fully in reverse. In particular, if, at any stage in the operation of a steam engine, some heat is allowed to fall from a higher temperature to a lower temperature without doing expansive work on the steam, that part of the operation will not be reversible. For example, if the vessel containing the heated steam is not well insulated and loses some of its heat by direct conduction to the cooler air that surrounds it, that part of the cycle will be irreversible, because by operating the engine backwards it would not be possible to recover the heat lost to the environment. A fully reversible steam engine is one in which no heat is ever allowed to flow directly from a hot body to a cold body or, equivalently, one in which the fall of heat is always mediated by the expansive work of the steam. Or, as Carnot put it, the condition of perfect reversibility and hence maximum efficiency in a heat engine is that *there should occur no changes of temperature, which are not due to changes of volume.* In practice, all actual working heat engines are irreversible to one degree or another.

Nevertheless, Carnot's theory is founded on consideration of the ideal, perfectly reversible heat engine. If, during one cycle, a reversible steam engine absorbs a certain quantity of heat from the furnace and produces a definite amount of expansive work, then by consuming that same amount of work we can operate the engine in reverse and restore to the furnace the same quantity of heat. That is Carnot's idea of reversibility.

Now, armed with these three great insights, namely, the fall of heat, the cycle, and the idea of reversibility, Carnot stated and proved an astonishingly general theorem regarding the operation of any and all heat engines. This theorem permitted him to answer the questions with which his *Reflections* began about the maximum efficiency possible in heat engines. As you will recall, engineers of the time wondered whether there was any assignable limit to the amount of work that could be extracted from a given quantity of heat and whether, in particular, an expansive agent other than steam might derive work from heat more efficiently.

What Carnot proved was, in his own words, that *the motive power of heat is independent of the agents employed to realize it; its quantity is fixed solely by the temperatures of the bodies between which the transport of caloric finally takes place*.

Therefore, *the maximum of motive power resulting from the employment of steam is also the maximum of motive power realizable by any means whatsoever.*

More precisely, what Carnot shows is that any reversible engine operating between two fixed temperatures produces work at the maximum efficiency possible for those temperatures; that is, no choice of a different working substance could possibly improve on the efficiency of a given reversible engine. His proof is a remarkably simple *reductio ad absurdum*. **[See p. 2 of the handout.]**

Suppose we have a working reversible heat engine that, after a certain period of operation, has absorbed a quantity of heat Q at the furnace while producing a quantity of work W. (Process A in the handout) Operated in reverse, our engine will consume a quantity of work W while discharging to the furnace the quantity of heat Q. (Process B)

Now imagine that there is a second, more efficient, engine that absorbs the same quantity of heat Q at the furnace while yielding a quantity of work $W+\Delta W$, greater than the work produced by our first engine. (Process C)

Suppose we now combine the forward operation of our second engine with the backwards operation of our first engine. (Process B+C) Our second engine absorbs heat Q at the furnace while yielding work $W+\Delta W$. (Process C) Our first engine, operated in reverse, will restore all the heat Q to the furnace while consuming only work W (Process B). The net result of B+C is that the quantity of work ΔW has been produced without taking any heat from the furnace.

At this point Carnot took a false step, guided by his belief that heat is indestructible. Since all the heat absorbed at the furnace has been returned to it by the end of our combined operation, Carnot assumed that the refrigerator has also given back all the heat it received. The net effect of the combined operation was, in his eyes, *the creation of work out of nothing*, and that is impossible, because it amounts to the operation of a perpetual motion machine.

But if we do not assume with Carnot that heat is indestructible and instead acknowledge that some of the heat is converted into work, we will have to conclude that, at the end of Carnot's combined operation, the work ΔW has been produced not out of nothing but out of heat extracted from the refrigerator. This would not be a perpetual motion machine in the original sense, but rather a machine that does work merely by extracting heat from the colder body.

Had Carnot remained agnostic on the question of the nature of heat, he could still have proved his theorem by basing the *reductio* on a different ground. Let me explain. **[See handout.]**

Consider again the three processes A, B, and C. Process B, you will recall, is the reverse operation of our Engine I, consuming work W while discharging heat Q to the furnace. Suppose the engine in this process is run a little longer, so that altogether it consumes work $W+\Delta W$ while discharging to the furnace a quantity of heat $Q+\Delta Q$. Let us call this prolonged version of Process B, Process D. If we now combine Processes D and C, the net result is that no work has been done, while a quantity of heat ΔQ has been transmitted to the furnace. Now both the caloric and the mechanical theory of heat agree in denying that heat can be produced out of nothing, when no work has been done. Therefore, on both theories, we are forced to conclude that the heat ΔQ given to the furnace has been extracted from the refrigerator, so the net result is that heat has been transferred from a cold body to a hot body without any other permanent change taking place. But that is impossible, according to Carnot, so our original assumption that Engine II is more efficient than the reversible Engine I must be false. *Q.E.D.*

Notice that Carnot's flawed *reductio* proof relied on the impossibility of producing work out of nothing, i.e., on the principle of the conservation of energy, which has come to be known as the First Law of Thermodynamics. In contrast, our modified version of his proof relies on the statement that *heat cannot flow spontaneously from a cold body to a hot body*. And this latter statement is the form in which Clausius first expressed his entropy principle, which we now call the Second Law of Thermodynamics.

It is often said that Carnot, with his insight about the fall of heat and his belief that heat was indestructible, discovered the Second Law of Thermodynamics without knowing the First Law. But that is quite misleading. As we can see in his rejection of perpetual motion, Carnot embraced the idea of energy conservation; he merely denied that heat is a form of energy. He believed that heat did work by falling down a temperature gradient, much as water does work by falling down a gradient of height. Carnot ruled out a spontaneous passage of heat from a cold body to a hot body on the same ground that he ruled out the spontaneous ascent of water up a hill; both would be violations of energy conservation, producing work out of nothing.

Ironically, then, for Carnot the impossibility of heat flowing from a colder to a hotter body is not an original and independent law but rather a routine consequence of the law of energy conservation. Because he misunderstood the nature of heat, he was unable to recognize the significance of what he had in fact found: the Second Law of Thermodynamics.

The theorem proved by Carnot has important consequences, both theoretical and practical. Practically, it means that a heat engine operating between two fixed temperatures can be made more efficient only by making its operations more perfectly reversible. Changing the working substance, say from steam to air or alcohol, will not raise the maximum rate at which work can be derived from heat.

Theoretically, Carnot's theorem means that, in a perfectly reversible heat engine, the work produced per unit of heat absorbed depends only the temperatures of the furnace and the refrigerator. Carnot himself never found the formula expressing that dependence, so he could not write down an expression for the efficiency of a reversible heat engine. That was one of the great tasks he left to his immediate successors, Lord Kelvin and Rudolph Clausius. More importantly, Carnot's work led, within a generation, to the discovery of entropy and the precise mathematical statement of the Second Law of Thermodynamics. To follow this development of Carnot's ideas, it will help to go a little further into his account of the operation of a reversible heat engine.

In a steam engine, the cycle usually involves a phase where the steam is cooled and compressed sufficiently to condense it into water. But that is not a necessary feature of the operation of heat engines. For simplicity's sake, we will assume a heat engine whose working substance is always in a gaseous state and is never liquefied. At any given moment, the gas occupies a certain volume and is assumed to have a definite uniform temperature and pressure. At this point, all we need to know about its behavior is that, like any gas, its temperature tends to fall spontaneously when it expands and to rise spontaneously when it is compressed, and, if the volume is held constant, the pressure will rise as the temperature rises and fall as the temperature falls.

Let us suppose a reversible heat engine operating between two fixed temperatures. That is, we assume that the furnace and refrigerator are always maintained at constant temperatures, so that the vessel containing gas always absorbs heat at a fixed higher temperature and discharges it at a fixed lower temperature. That will be the case if the furnace and refrigerator are assumed to be so large compared to the vessel of gas that their temperatures are practically unaffected by the operation of our engine. Following Carnot, we shall call the furnace body A and its temperature T_A; the refrigerator is body B and its temperature is T_B.

The cycle that Carnot described has four phases. **[See p. 3 of the handout.]** At the start of the cycle the vessel of gas is assumed to be at the higher temperature T_A. It is put into contact with the furnace, body A, and allowed to expand while maintaining the same temperature T_A. Since its natural tendency is to fall in temperature, it must absorb heat from body A in order to maintain a constant temperature. Since the gas is expanding, it does useful work in driving the piston. Because its expansion all takes place at constant temperature, we shall call this the phase of **isothermal expansion**.

We now remove the gas from contact with body A but allow it to continue expanding and thus to continue doing work by driving the piston. The temperature of the gas will fall spontaneously during this further expansion, but the gas neither gains nor loses any heat from its surroundings. A process during which no heat passes into or out of a body is called an **adiabatic** process, from the ancient Greek word *adiabatos*, which means "not to be passed." Incidentally, this word makes its first appearance in Xenophon's *Anabasis* where it describes certain wild and impassable Babylonian rivers. Be that as it may, we shall call this second phase of the cycle the phase of **adiabatic expansion**, during which no heat is either absorbed or expelled. This phase ends when the gas has expanded and cooled to the point where it has fallen to the temperature T_B of body B, the refrigerator. At this point it is safe to bring the gas into contact with body B; since they are now at the same temperature, no heat will be passing from a hot body directly to a cold body; in Carnot's terms, there will be no "useless restoration of the equilibrium of caloric," and the process will remain reversible.

During the third phase, we compress the gas while maintaining it at temperature T_B. Since its temperature tends to rise spontaneously during this compression, it must discharge heat to body B in order to maintain its low temperature. To compress the gas during this phase, work must be done *on* the engine. This third phase is called the phase of **isothermal compression**.

Finally, for the fourth phase we remove the gas from contact with body B and continue compressing it without passage of heat to or from the gas. The gas spontaneously rises in temperature during this phase of **adiabatic compression**. The compression, during which work must be done *on* the engine, continues until the gas has risen to T_A, the temperature of the furnace. At that point, it is once again safe to bring it into contact with the furnace, body A, and to begin another period of isothermal expansion.

Whatever state of volume and pressure the gas is in when it reaches temperature T_A at the end of the adiabatic compression will thereafter be considered the starting point of every subsequent cycle. Henceforth, at the end of each cycle the gas has been restored to exactly the same temperature, pressure, and volume that it had at the beginning of the cycle. We can therefore be sure that the only effect of the cycle is to do expansive work while taking in heat at the higher temperature and discharging heat at the lower temperature.

Now you may have been disturbed to learn that the engine does work only during the two expansive phases, while work must be done on the engine during the two phases of compression. Only if the expansive work exceeds the compressive work will the cycle as a whole produce usable work. Fortunately, that is the case, because the changes of volume during expansion all occur at higher temperatures, and therefore at higher pressures, than the corresponding changes of volume during compression. Accordingly, in a reversible heat engine the expansive work done by the engine always exceeds the compressive work done on the engine, so the net effect is that the engine does a certain quantity of external work for each cycle of its operation.

For brevity's sake, we shall call the two phases of isothermal compression and expansion **isotherms**, and the two phases of adiabatic expansion and compression we shall call **adiabats**. The cycle we have just described, in which a reversible heat engine executes two isotherms alternating with two adiabats, is what has come to be known as a **Carnot cycle**.

Carnot cycles are very easy to visualize with the aid of a kind of diagram introduced by the French engineer Émile Clapeyron in 1834. We shall use Clapeyron's diagrams in a slightly modified form. In Cartesian coordinates we plot temperature on the vertical axis and volume on the horizontal axis. A point in the quadrant then represents a particular momentary state of the gas undergoing a Carnot cycle, and a closed curve in the quadrant represents a cyclical process, i.e., one which returns the gas to its original state. Isothermal processes are represented in the diagram by horizontal lines, since they occur at constant temperature. Moving from left to right on a horizontal line means increasing volume (i.e., expanding) at constant temperature. In our diagram, the engine cycles clockwise from point 1 to 2 to 3 to 4 and back to 1 again. Between points 1 and 2 the gas undergoes isothermal expansion while in contact with the furnace; between 2 to 3 it expands adiabatically while cooling down to temperature T_B. From 3 to 4 it is compressed isothermally while in contact with the refrigerator. And from 4 back to 1 the gas undergoes adiabatic compression until it again reaches the temperature of the furnace. Note that the gas gains heat on the isotherm from 1 to 2, loses heat on the isotherm from 3 to 4, and neither gains nor loses heat on the adiabats 2-3 and 4-1. Finally, work is done *by* the engine during the expansive part of the cycle from 1 to 3; work must be done *on* the engine during the compressive part of the cycle from 3 back to 1.

Since the Carnot cycle is perfectly reversible, let us see what happens when it is run backwards, cycling through the points of our diagram in reverse order, i.e., counterclockwise. Suppose the reversal takes place at point 2, the end of the isothermal expansion phase; we keep the gas in contact with the furnace, body A, but now we compress it isothermally, while discharging heat to body A at temperature T_A; when it reaches point 1 in the diagram, we remove it from body A and allow it to expand adiabatically until its temperature falls to T_B, point 4 in the diagram; then we put it in contact with body B and allow it to continue expanding isothermally while absorbing heat from body B, and finally (at point 3) we remove it from body B and compress it adiabatically until its temperature rises to T_A (point 2).

By operating the heat engine backwards for one cycle, we entirely neutralize the effects of the original cycle. Whatever work was done by the engine in the original cycle must be consumed by the engine in the backwards cycle. Whatever heat was absorbed from the furnace is now discharged to the furnace, and whatever heat was discharged to the refrigerator is now absorbed from the refrigerator. Operating in reverse, the work of compression exceeds the work of expansion since the compression takes place at higher temperatures. The backwards operation of the heat engine is a process of refrigeration, in that we are doing work on the machine, with the result that heat is extracted from a cold body and given to a hot body.

For the rest of the nineteenth century, despite his embrace of the ill-starred caloric theory, Carnot's idea of the perfect heat engine as a reversible cyclic process operating between two temperatures became the indispensable vehicle for all serious progress in thermodynamics. It captured the imagination of all the great scientists who came across Carnot's work, and it led more or less directly to the completion of classical thermodynamics in the 1850s and 60s.

To bring our discussion of the Carnot cycle to a close, we must mention once more the awkward fact that Carnot's brilliant achievements in thermodynamics were tangled up with an erroneous theory of heat, the caloric theory, according to which heat can neither be destroyed nor created. Carnot believed, incorrectly, that all the heat absorbed at the furnace must also be discharged at the refrigerator. We now know, however, that some of that heat is converted into work, so that the heat discharged to the refrigerator must be less than the heat absorbed at the furnace. How it is that Carnot's adherence to the caloric theory did not undermine the validity of his main results is a fascinating story well beyond the scope of our present discussion. In the meantime, as we move forward in our history, let us remind ourselves of Carnot's three key insights which he bequeathed to the next generation: the fall of heat, the cycle, and the idea of reversibility.

III. Clausius

For a generation after it appeared in 1824, Carnot's book was almost entirely neglected. It was then rediscovered by Kelvin and Clausius around 1850. In the meantime, the investigations of James Joule had undermined the authority of the caloric theory of heat. Kelvin himself, however, still believed in the conservation of heat when he embraced Carnot's ideas in 1848, and it seemed to him highly doubtful that Carnot's treatment of the heat engine could survive if one abandoned the axiom that heat is indestructible. It was Clausius, in 1850, who first saw clearly that Carnot's main ideas could be salvaged even if the caloric theory of heat was discarded in favor of the mechanical theory. Clausius showed how one could give up Carnot's assumption that no heat is lost in the operation of the Carnot cycle, while modifying Carnot's *reductio* proof of the theorem that a reversible heat engine attains the maximum possible efficiency. Carnot's proof then remains essentially valid, with the correction that we have already discussed.

What Clausius attempted was to combine the mechanical theory of heat with what he regarded as the sound part of Carnot's insight regarding the fall of heat. That is, he tried to ground the science of thermodynamics in two fundamental principles or laws. According to the first principle, energy is conserved, while heat and work, being forms of energy, are uniformly interconvertible. Clausius had a harder time formulating the second principle with precision. One of his early formulations is *that heat cannot of itself pass from a colder body to a warmer body.* Another formulation is that *heat has a spontaneous tendency to pass from a warmer body to a colder body.* Clausius's great achievement in the 1850s was to translate this principle into a precise mathematical statement.

According to Clausius, Carnot "regarded the production of work as the equivalent of a mere transport of heat from a hot body to a cold one, the quantity of heat being thereby undiminished." Clausius wants to discard that last clause while retaining the rest of the principle. He writes,

> although we have no need of a peculiar equivalent for the produced work, after we have assumed as such an actual consumption of heat, it is nevertheless possible that the said transport of heat may take place contemporaneously with the consumption, and may likewise stand in a certain definite relation to the produced work.

In other words, both the consumption of heat and the transport of heat from a hot to a cold body may be necessary conditions for the production of work. But it is not yet clear what definite relation the transport of heat has to the production of work. Clausius was to solve this problem in his next great contribution to thermodynamics, the paper of 1854.

In that paper Clausius presents a modified form of the Carnot cycle designed to separate out the two conditions for the production of work, namely, the consumption of heat and the fall or transport of heat from a hot body to a cold one.

Clausius's complex Carnot cycle operates at three different temperatures instead of two. In effect, he has two furnaces and one refrigerator. **[See p. 4 of the handout.]** Referring to the diagram, the cycle begins at point 1, where the gas, in contact with furnace A, begins isothermal expansion at temperature T_A. At point 2 it is removed from furnace A and allowed to expand and cool adiabatically to temperature T_B, which it reaches at point 3. From 3 to 4 it expands isothermally while in contact with furnace B at temperature T_B. Between 4 to 5 it expands adiabatically and cools down to temperature T_C. From 5 to 6 it is compressed isothermally while in contact with refrigerator C, and from 6 to 1 it is compressed adiabatically until it has warmed up to temperature T_A. In effect, the gas absorbs heat successively from two furnaces at two different temperatures after which it discharges heat to the refrigerator at a third temperature.

Let us remind ourselves that, in Carnot's original presentation of the simple 4-point cycle, he believed that no heat was consumed, so that the heat absorbed from the furnace was equal to the heat discharged to the refrigerator. But Clausius's principle of the interconvertibility of heat and work requires that the heat gained during isothermal expansion at the furnace exceed the heat lost during isothermal compression at the refrigerator. The excess is the net heat converted into work during one cycle. In our diagram of the simple Carnot cycle, that means that more heat is absorbed between points 1 and 2 than is discharged between points 3 and 4. Similarly, in Clausius's complex 6-point Carnot cycle, the heat absorbed from the two furnaces exceeds the heat lost to the refrigerator. Clausius is therefore free to adjust the durations of the two isothermal expansion phases, phases 1-2 and 3-4, so that the heat absorbed between points 3 and 4 exactly equals the heat lost between points 5 and 6. Let us call each of these quantities of heat Q.

Since the heat gained between 3 and 4 is cancelled out by the heat lost between 5 and 6, and since no heat is gained or lost on the adiabats 2-3, 4-5, and 6-1, we conclude that, from point 2 clockwise all the way around to point 1, no heat is gained or lost. Therefore, the only heat gained in the whole cycle must be gained between points 1 and 2, i.e., during the first isothermal expansion. But since, at the end of the cycle, the gas has returned to its original state and has neither gained nor lost any heat, the heat gained between 1 and 2 must have been entirely converted into work. Let us call this quantity of heat Q_1.

In this way, Clausius has constructed a cycle in which a certain quantity of heat Q_1 has been entirely converted into work at the high temperature T_A, while another quantity of heat Q has been allowed to fall from the intermediate temperature T_B to the lower temperature T_C. He has thus broken down the Carnot cycle into two different kinds of transformation, the first kind being a conversion of heat into work, the second kind being a fall of heat in the original Carnotian sense.

What is the relation between these two different transformations? Clausius's surprising answer is that they are *equivalent* to one another. But how is it possible for a conversion of heat into work to be equivalent to a fall of heat from one temperature to another? Referring again to the diagram of the complex cycle, we note that the conversion of heat Q_1 into work on the first isotherm is a way of passing from point 1 to point 2 on the cycle. To get from point 2 back to point 1 one could simply follow the first isotherm backwards, i.e., as a compression. But one could also continue clockwise from point 2 all the way around the cycle to point 1, allowing heat Q to fall from T_B to T_C. In effect, either of the two transformations, when reversed, can take the place of the other transformation in the cycle. It is for that reason that Clausius calls the two transformations equivalent.

Clausius now takes a very bold mathematical leap. He conjectures that it should be possible to assign a mathematical magnitude to each transformation, in such a way that equivalent transformations have equal magnitudes. He calls the magnitude of a given transformation its *equivalence-value.* He then sets out to find the law that will assign appropriate equivalence-values to all transformations. Through an ingenious argument whose steps we cannot follow here, Clausius concludes that the equivalence value of an isothermal conversion of work into heat is equal to $\frac{Q}{T}$, where Q is the heat and T is the temperature. The reverse operation, the conversion of heat into work, has the equivalence value $-\frac{Q}{T}$. As for a fall of heat from temperature T_1 to temperature T_2, Clausius argues that the equivalence value assigned to it should be $Q\left(\frac{1}{T_2}-\frac{1}{T_1}\right)$. **[See handout.]**

By comparing the two expressions, Clausius shows that the second kind of transformation, the fall of heat, is also equivalent to a double transformation of the first kind, i.e., a conversion of heat into work at the higher temperature and back from work into heat at the lower temperature. Indeed, Carnot draws the general conclusion that every transport of heat from one body to another is equivalent to a conversion of heat into work and back again into heat.

On this basis Clausius shows how to assign a total net value to all the transformations of both kinds in any cyclical process, however complicated. One merely looks at each quantity of heat lost or gained by any body at any temperature and sums up the values of $\frac{Q}{T}$ or $-\frac{Q}{T}$ for all these processes.

In the case of Clausius's six-point Carnot cycle, the fall of heat has the same equivalence value, though with opposite sign, as the conversion of heat into work. Therefore, traveling all the way around the cycle, the sum of the equivalence values will be zero. Clausius shows in general that every reversible cyclical process must have a net equivalence value of zero. In the six-point cycle, for example, the fall of heat has a positive equivalence-value which is exactly compensated by the negative equivalence-value of the conversion of heat into work.

In a simple 4-point Carnot cycle, where the conversion of heat is not kept separate from the fall of heat, it is clear that the equivalence values of the two isotherms must be equal and opposite, In other words, the value $\frac{Q}{T}$ for the heat absorbed at the higher temperature, must equal that of $\frac{Q}{T}$ for the heat expelled at the lower temperature. Since the difference between the heat absorbed and the heat expelled is the heat converted into work on one cycle, Clausius was immediately able to write down the formula for the efficiency of a Carnot cycle operating between any two temperatures, a feat that had eluded Carnot.[3]

It took him another eleven years to see it, but what Clausius had discovered here in his 1854 paper was a new fundamental quantity in nature, the entropy. The quantity of heat contained by a body can change either by direct conduction of heat to or from another body or by conversion of its heat into work. In either case, a change in the heat content of a body is always accompanied by a change of entropy. Whenever heat enters a

---

[3] If Q is the heat absorbed and J is the work produced per unit of heat converted, then the efficiency of the perfect (reversible) heat engine is given by $\frac{W}{JQ} = \left(\frac{T_1 - T_2}{T_1}\right)$.

body at a given temperature the body's entropy is increased by an amount equal to $\frac{Q}{T}$. Whenever heat leaves a body, its entropy decreases by $\frac{Q}{T}$. Now consider Clausius's two kinds of transformations. When heat is converted into work, the entropy decreases, whereas when heat falls from a hot body to a cold body, the net entropy in both bodies increases. When these two operations are combined in a Carnot cycle, the total change in entropy for all bodies is zero, since the production of work is exactly compensated by the fall of heat, and the fall of heat by the production of work.

Recall that, in a reversible cycle, heat is never allowed to fall from a hot body to a cold body without doing work by means of a heat engine. In Clausius's new way of speaking, there must be no uncompensated falls of heat if a cycle is to be perfectly reversible. In particular, heat can never be converted into work without a compensating fall of heat from a hot to a cold body, and heat can never flow from a cold to a hot body without a compensating consumption of work. One recognizes in these statements Clausius's early qualitative expressions of the Second Law of Thermodynamics.

What about an irreversible cycle? We know that it is always possible for heat to flow directly from a hot body to a cold body without the intervention of a heat engine. In that case, the increase of entropy due to the fall of heat is not compensated by the decrease of entropy due to the production of work. Therefore, in irreversible cycles, the net entropy must always increase.

But since, in the real world, no machine ever operates with perfect reversibility and there are always some uncompensated flows of heat from hot to cold bodies, one may conclude that, in every actual cyclic process, the entropy change must always be greater than or equal to zero, with perfectly reversible processes forming the lower limit. And that is Clausius's mature, quantitative statement of the Second Law of Thermodynamics.

Looking backwards one may see that Clausius's achievement is in a real sense a vindication of Carnot's insights regarding what goes on in heat engines. Carnot thought of heat at a higher temperature as somehow more energetic, more capable of doing work, than heat at a lower temperature. Accordingly, to allow heat to pass directly from a hot body to a cold one is simply a waste of available energy, uncompensated by useful work. Since the measure of entropy change is heat lost or gained divided by temperature, we can now see that high temperature heat is low entropy heat, and vice versa. The uncompensated flow of heat from hot to cold bodies is thus a spontaneous increase of entropy, and a permanent loss of some portion of the usable energy in the universe. Not only did Carnot see that a fall of heat is necessary in order for heat to do work, he also believed that something absorbed from the furnace is passed undiminished to the refrigerator. He thought it was the heat, and he was wrong. But it does turn out that, in a reversible engine, exactly the same amount of entropy that enters the engine at the furnace also leaves the engine at the refrigerator.

IV. Implications of the Discovery of Entropy

I have already reported the gloomy expectations that the Second Law of Thermodynamics has inspired in some authors. In contrast to classical Newtonian mechanics, thermodynamics teaches that, for any closed system—and for the universe as a whole—the future is essentially different from the past, and probably not for the better. Whether such thinking justifies a general mood of despondency or rather a resolute determination to eat, drink, and be merry, is perhaps a question we might discuss. But to bring this talk to a timely end, I would like to briefly mention two other implications of the discovery of entropy that seem worth pondering.

First, there is the perennial question of how much our understanding of the human things and of our place in the whole can be affected by the discoveries of modern science. In this case, that question takes an especially interesting turn. I think there is little doubt that the existence and nature of entropy could not have been discovered apart from the modern scientific project. And it is well known that that project, at its origin, is closely associated with the goal of the mastery of nature for the enlargement of human power. But in this case, we find the peculiar circumstance that a major theoretical advance was achieved—and, I would argue, could only have been achieved—by intense reflection on a nitty-gritty problem of engineering, specifically, how to improve the efficiency of a steam engine. Indeed, the science of thermodynamics was discovered not so much through contemplation of nature, but through study of machinery. The Carnot cycle is after all an idealization of the operation of a man-made *engine*, not of anything that is ever encountered in nature. The discovery of entropy, then, seems to be a case where not only modern science, but even modern technology as such—grimy, sooty, and oil-soaked—has produced key theoretical insights into the nature of the world and our place in it. This seems to me a point worth discussing.

Second, there is a subtle and interesting question concerning the theoretical status of the two Laws of Thermodynamics themselves. When we label something a *law of nature*, does that grandiose phrase mean anything more than a generalization from experience? Interestingly, Albert Einstein, who was typically skeptical of any physical theory's claim to absolute truth, wrote of the "deep impression" made upon him by classical thermodynamics: "It is the only physical theory of universal content concerning which I am convinced that, within the framework of the applicability of its basic concepts, it will never be overthrown."[4] As I mentioned earlier, the conservation of energy, the subject of the First Law, seems to most people a more clear and intuitive idea than the increase of entropy, the subject of the Second Law. The former would appear to rest on the axiom, or intuition, that a perpetual motion machine is impossible, i.e., that work cannot be done without the expenditure of energy. This is but a more precise expression of the age-old conviction that there can be no science of nature—in fact, there can be no nature at all—unless "nothing can come out of nothing." That is a conviction that was shared by all the Greek philosophers, from the atomists to the idealists, and perhaps by every scientist since. But what is its basis? The scientific champions of the First Law were inclined to invoke theological considerations. Joule, for example, declared that "the grand agents of nature are, by the Creator's fiat, *indestructible*; and wherever mechanical force is expended, an exact equivalent of heat is *always* obtained"[5] (emphasis in original); and Kelvin tells us he is certain that "Creative Power alone can either call into existence or annihilate mechanical energy."[6] Meanwhile, in the Book of Exodus, when God wanted to attract the attention of Moses, he set in his path (Exodus 3.2) "a bush that burned with fire but was not consumed." Only when Moses turned aside to see for himself why the bush burned but did not burn up did God reveal his plan for the liberation of the children

---

[4] Albert Einstein, "Autobiographical Notes," in Paul A. Schilpp, ed., *Albert Einstein*: *Philosopher-Scientist* (Evanston, Illinois: Library of Living Philosophers, 1949), p. 33.

[5] James Prescott Joule, "On the calorific effects of magneto-electricity, and on the mechanical value of heat," *Philosophical Magazine*, Series 3, 23: 263–276 (1843).

[6] Sir William Thomson (Lord Kelvin), "On a Universal Tendency in Nature to the Dissipation of Mechanical Energy," *Proceedings of the Royal Society of Edinburgh* for April 19, 1852; *Mathematical and Physical Papers*, vol. 1, art. 59, p. 511.]

of Israel from slavery. Evidently the qualities God was looking for in the leader of the Jewish people were scientific curiosity plus an intuitive grasp of the First Law of Thermodynamics.

And yet, an intuition is not a proof, and the ultimate ground for our belief in the conservation of energy remains an enigma. Perhaps the only significant advance on this question was made in 1915 by Emmy Noether, the greatest woman mathematician in history. What she proved was that every conservation law is the consequence of a particular symmetry that is found in nature. The conservation of energy in particular is associated with symmetry or invariance of a certain dynamical function, called the Lagrangian, with respect to translation in time. Of course, Noether's Theorem only shifts the question of the ground of energy conservation to the equally perplexing question, why are there such symmetries in nature?

As for the Second Law of Thermodynamics, it appears to rest on the intuition, or axiom, that heat cannot *spontaneously* flow from a cold body to a hot body, i.e., that it can be made to flow in that direction only if work is done on the system. I leave it to you to judge whether this rule is anything more than a generalization from experience.

However, only a few years after Clausius coined the term "entropy" in 1865, the Austrian physicist Ludwig van Boltzmann began a program of research on the random molecular motions that underlie the large-scale phenomena studied by thermodynamics. What Boltzmann created was the science of "statistical mechanics," one of whose principal results is a new and deeper understanding of the entropy law. According to Boltzmann, every macroscopic state of a body (for example, a volume of gas at a given temperature) has available to it a multitude of possible micro-states (in our example, the innumerable ways that mechanical energy can be distributed among the molecules of a gas in that state). Boltzmann showed that the entropy of a system is a function of the number of micro-states available to a given macro-state. And as all micro-states are presumed to be equally probable, the Boltzmann entropy turns out to be a function of the relative *probability* of the given macro-state. Translated into statistical mechanics, the Second Law of Thermodynamics expresses the fact that, over time, systems composed of many particles in random motion tend to evolve from less probable states to more probable states, and not vice versa. (Sometimes, the Second Law is informally said to involve a tendency toward greater *disorder*; but that is only because ordered states are, for the most part, highly *improbable*.) Heat tends to flow from a hot body to a cold body simply because a collection of particles in random motion yet with all the hotter (i.e., faster-moving) particles in one place and all the colder ones in another is highly unlikely to arise by chance, whereas a more or less uniform mixture of faster and slower molecules is much more likely. In the same way, a box of coins all facing heads up is likely, when shaken, to become a box with about half of the coins facing down. Further shaking is unlikely to reproduce the original improbable configuration with all coins heads up. If you find this intelligible, you have grasped the essence of the Second Law of Thermodynamics as re-interpreted by Boltzmann. Hence my suggestion that, contrary to initial appearances, it is the First Law of Thermodynamics, the conservation of energy, that is ultimately mysterious as to its ground, whereas the Second Law, the increase of entropy, at least as clarified by Boltzmann, is the epitome of logic and common sense.