

The Use and Misuse of the Word "Heat" in Physics Teaching

Mark W. Zemansky

Introduction

In the period from 1925 to 1966, the author has tried to teach the subject of thermodynamics to students who have been almost exclusively undergraduates in regularly scheduled American college classes. Beside this experience, the author has been called upon to lecture to high-school teachers and high-school students on various topics in thermal physics at summer institutes sponsored by the National Science Foundation and by the Visiting Scientist Program of the American Association of Physics Teachers. During this period the author has been struck by the confusion existing in the minds of those whose main source of information consisted of elementary textbooks of physics and chemistry. The well-known mixup between the words temperature and heat he ascribes to the newspapers who habitually refer to a "heat of 90 degrees." This error is quite serious, but seems hardly worthwhile discussing at this point.

The present paper is the result of the perusal of books at the elementary level written by two classes of authors. On the one hand, there are high-school teachers with enormous experience in teaching young people, but with only a scant knowledge of thermal physics; and on the other hand, there are college professors with considerable knowledge of thermal physics and with great research background, but with only slight experience in the teaching of elementary students. It is an interesting commentary that both kinds of author make the same mistakes, one out of a mistaken notion of what thermal physics is, and the other from a mistaken notion of what simplifica-

tion implies. I should like to set the record straight for the unwary high-school teacher-author, and for the equally unwary college professor-author. In doing this, I shall make use of some quotations only from books written by college professors. But I shall not disclose the titles nor the authors, for my purpose is not to criticize people, but to ridicule the point of view expressed by the jingle:

Teaching thermal physics
Is as easy as a song:
You think you make it simpler
When you make it slightly wrong!

The First Law of Thermodynamics

In some respects, the first law of thermodynamics is more subtle and more difficult to grasp than the second law. The first law involves the concept of internal energy, whose meaning many people feel is obvious and whose name they think may be replaced by other words such as "heat" or "thermal energy." We have heard the word energy so often that we think we "feel" its meaning and can do without a definition. Furthermore, there is no definition of energy in general, and some people do not like to believe that the "internal energy" of the first law of thermodynamics has a very special definition and meaning, bringing with it a very special meaning to the word "heat."

Before the first law can be established, it is necessary to define two kinds of walls that are generalizations of walls that may be constructed easily. When an *adiabatic* wall is used to separate two systems, *A* and *B*, the macroscopic properties of one system are found to be uninfluenced by those of the other. For example, if a concrete wall $\frac{1}{2}$ m thick separates a coil of resistance wire from the hot gas issuing from a furnace, neither the properties of the wire (tension, length, electric resistance) nor the properties of the gas

Paper delivered at a meeting of "Groupe International de Recherche sur L'Enseignement de la Physique," held in Copenhagen, 1969. This paper, along with the others presented in Copenhagen, is published in the GIREP "Proceedings" by Gyldendalske Forlag, Klareboderne 3, Copenhagen, Denmark.

(pressure, volume) are affected. Should the wall, however, be a thin sheet of copper (*diathermic wall*), the two systems will undergo changes until *thermal equilibrium* is reached. It is a simple matter to show how the concept of temperature arises from considerations of this sort, and for further details the reader is referred to any textbook of thermodynamics.

The first law of thermodynamics starts with the concept of work, which always means the same thing in all branches of science. If there exists a macroscopic force F , undergoing a macroscopic displacement ds in a direction parallel to the force, the work dW is defined as Fds . Each simple system has its own expression for work, such as PdV for a hydrostatic system, $-HdM$ for a magnetic system, etc. Sometimes two or more simple expressions must be used to express the work of a more complicated system, such as a paramagnetic gas whose work dW is given by $PdV - HdM$. Simple calculations show that the work in a finite process depends on the set of processes (path) used to change the system from its initial to its final state. There is no such thing as the "work in a system." The work is not a function of the coordinates of a system.

In Fig. 1(a) there are depicted two ways in which *adiabatic work* may be done on a system. Under the action of force F , motion may be imparted to a piston; and with the aid of a falling object an electric generator may be rotated, thereby providing a current in a resistor embedded in the system and included as part of the system. If we focus our attention on two arbitrarily chosen states of the system i and f (initial and final), it is possible, with the aid of the apparatus of Fig. 1(a), to proceed from i to f along many different paths. For example, one could first allow the electric current to exist for a minute and then push the piston in or out; or one could move the piston and then maintain the current. No one, as far as this author is aware, has ever measured the adiabatic work performed in a variety of adiabatic

paths joining two arbitrary states of a system. Nevertheless, it is accepted as a fact (indeed as a law of nature, which is often called the first law of thermodynamics), that the adiabatic work is independent of the path.

It is a fact of experimental physics that the apparatus depicted in Fig. 1(b), in which an adiabatic wall is replaced by two metal walls whose temperatures are maintained equal at all times, is in all respects equivalent to an adiabatic wall. This temperature equality is achieved experimentally by noting the reading of a differential thermocouple whose junctions touch the surface of the system and the surrounding wall, and by supplying energy to either the system or the surrounding wall whenever this reading differs from zero.

Using this equipment, it is possible to measure the work W_{ad} when the system is brought from state i to state f adiabatically (when both metal walls are at the same temperature), and to observe that W_{ad} is independent of the path. The consequences of this independence of path may now be derived. Going back to mechanics, the student may be reminded that the work to move an object slowly and without friction from the floor to a point on a desk is independent of the path. This was shown to lead to the concept of gravitational potential energy. Further discussion of the way in which kinetic energy and elastic potential energy were introduced into elementary physics may be utilized to motivate the conclusion that, since the adiabatic work in bringing a system from state i to state f is independent of the path, there must exist a function U of the coordinates of the system, such that

$$\Delta U = U_f - U_i = -W_{ad}.$$

The minus sign ensures that work done *on* a system (negative work) makes U_f larger than U_i . The function U could be called the "adiabatic work function" quite justifiedly, but is called instead the *internal energy function*.



Mark W. Zemansky served on the physics faculty of the City College of New York for forty-six years, three of them as department chairman and another three as Executive Officer of the physics doctoral program of the City University. He has been president and Oersted medal winner of the American Association of Physics Teachers and is now its Executive Secretary.

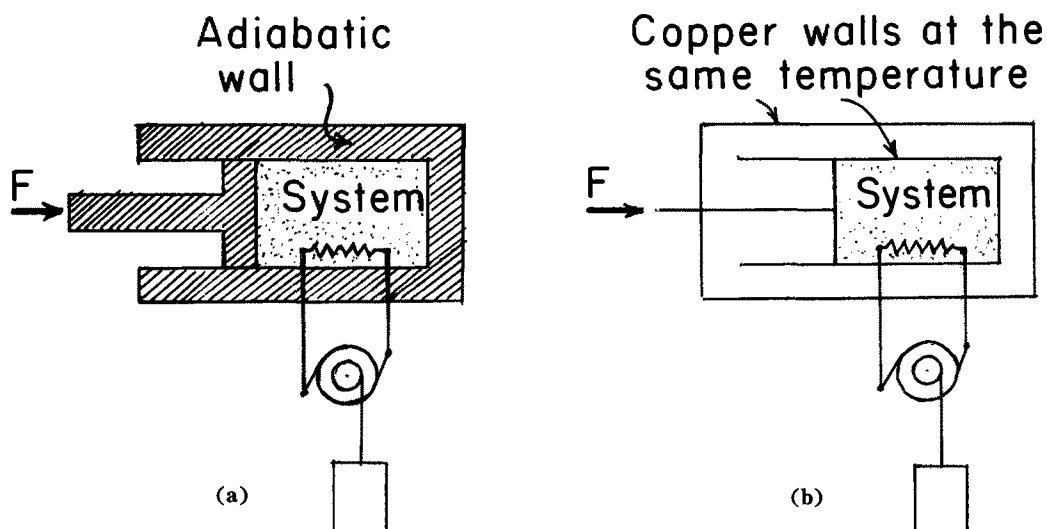


Fig. 1. The behavior of the system surrounded by adiabatic walls in (a) is the same as that whose walls and surroundings are at the same temperature in (b).

Going back to Fig. 1(b), suppose that the outer metal wall is deliberately set at a temperature different from that of the inner wall, and the system is now brought from i to f and the work W is measured. This work would be found to be different from W_{ad} , that is, different from $-(U_f - U_i)$. Since it is undesirable to give up the principle of the conservation of energy, the conclusion is inescapable that the difference between W and W_{ad} [or W and $-(U_f - U_i)$] is due to another form of energy that must have entered or left the system *by virtue of the temperature difference between the system and its surroundings*. This energy flow, brought about by a temperature difference, is none other than the quantity that has always been called "heat." Long before the science of thermodynamics was established, experiments in which hot and cold liquids were mixed, or hot pieces of metal were immersed in a beaker of cool water, etc., had given rise to the subject of "calorimetry" and had provided the *calorimetric* definition of heat as the energy transferred by virtue of a temperature difference. We see, therefore, that this old calorimetric definition and the more modern and more exact thermodynamic definition are in agreement. Denoting the heat that has entered or left the system by the symbol Q , we get

$$Q = W - W_{ad} = W + U_f - U_i,$$

which is taken today to be the *mathematical formulation of the first law of thermodynamics*.

Rewriting the first law in the form

$$U_f - U_i = Q - W$$

shows that Q and W are *methods of producing a change in internal energy*. W is the mechanical method of energy transfer, and Q is the non-mechanical method. Together they constitute the

two thermodynamic methods in which U may be changed.

It has been emphasized that W , being nonadiabatic, depends on the path and is therefore not a function of the coordinates of the system. There is no such thing as the "work in a body." For identical reasons, Q also depends on the path and is not a function of the system's coordinates. There is no such thing as the "heat in a body." To quote from Slater's *Introduction to Chemical Physics*:

"At first sight, it seems too bad that Q is not independent of the path, for some such quantity would be useful. It would be pleasant to be able to say, in a given state of the system, that the system has so and so much heat energy. Starting from the absolute zero of temperature, where we could say that the heat energy was zero, we could heat the body up to the state we were interested in, find Q from absolute zero up to this state, and call that the heat energy. But the stubborn fact remains that we should get different answers if we heated it up in different ways. For instance, we might heat it at an arbitrary constant pressure until we reached the desired temperature, then adjust the pressure at constant temperature to the desired value; or we might raise it first to the desired pressure, then heat it at that pressure to the final temperature; or many other equally simple processes. Each would give a different answer, as we can easily verify. There is nothing to do about it."

Heat and work are methods of energy transfer, and when all flow is over, the words heat and work have no longer any usefulness or meaning. As a result of these transfers, the internal energy of the system has either increased, decreased, or remained constant, and once the transfers are over, we can speak only of the internal energy of the system. It is impossible to subdivide the

internal energy into two parts, one due to a heat transfer and the other to work, just as it is impossible to subdivide the water in a lake into parts resulting from (1) ground seepage, (2) rain, (3) flow from a brook, (4) condensation, etc.

Sometimes the decision as to whether an energy transfer has been accomplished by the performance of work or by a flow of heat is a bit fussy. Go back for a moment to Fig. 1(a) and imagine the material marked "system" is water. Suppose that the piston remains stationary and that a current (produced by the falling object) exists for 1 min. There are three choices as to what constitutes the "system":

(1) *The water alone.* As soon as the resistance wire becomes a trifle hotter than the water, heat flows into the water. No work is done *on the water alone*, so that

$$W = 0, \quad Q = \Delta U \text{ (of the water).}$$

(2) *The wire alone.* Work is done on the resistor, and heat flows out of the resistor without producing any net change in the coordinates of the wire. Hence

$$\Delta U \text{ (of the wire)} = 0; \quad Q = W.$$

(3) *Water and wire.* The process is adiabatic, so that

$$Q = 0, \quad W = \Delta U \text{ (of the water and wire).}$$

Three Commonly Made Errors

There are three main infelicities of expression that are indulged in by writers who are trying to come down to the level of introductory physics or chemistry. They are

- (1) Referring to the "heat in a body."
- (2) Using "heat" as a verb.
- (3) Combining heat and internal energy into one undefined concept "thermal energy," which on one page means heat and on the next page means internal energy.

(1) *The "heat" in a body.* Consider the following quotations:

"The kinetic theory of heat leaves no doubt that what we usually call 'heat' is the manifestation of the internal micromotion of the molecules that form all material bodies, and that 'temperature' characterizes the intensity of the internal molecular agitation."

"The word *temperature* means the intensity of heat and may be defined as a number on a scale."

"Energy becomes heat as soon as it is disordered. It is conversely true that disorder can exist without energy, and that disorder becomes heat as soon as it is energized. To see how heat is produced by adding energy to disorder, consider the air in a bicycle pump. Before compression, the air atoms are already moving at random in all directions; in other words, this is a disordered

system and its energy is in the form of heat. Now, if you pump vigorously, compressing the air rapidly, it heats up; the pump becomes hot to the touch. The air has the same disorder it had before, but more energy. By doing work you have pushed more energy into the air, and the observed production of heat is just the effect of this addition of energy to the pre-existing disorder."

"The small amount of hot water has a higher temperature, but a lower heat content than the much larger amount of cool water. The hot water contains more heat than an identical volume of cool water."

"The energy of a gas is the energy of molecular motion, and we call this its *heat energy*. You may say 'Heat is molecular motion;' but please avoid saying 'Heat is due to collisions among molecules.' That is entirely wrong. During a collision, one molecule may give some heat energy to another; but collisions cannot manufacture heat, and they certainly are not heat. If the moving molecules of a gas suddenly shrank, so that most hits became misses and they made far fewer collisions, they would still have the same heat energy at the same speed. . . . When gas molecules hit a *moving* piston, they rebound with different speed—faster if the piston was approaching them. Such molecules do gain heat when they hit the piston, but they gain it from the piston's energy, not from some mysterious spark called a collision."

"Now, although ice has a 'rigid' crystalline form, its temperature can change—ice has heat. If we wish, we can change the amount of heat. What is the heat in the case of ice?"

"We shall find in our later work that *temperature* is a measure of the *average translational energy* per molecule and that *heat* or heat content is a measure of the *total* molecular energy."

"If a quart of water at 50°C is separated into two pints, each pint remains at the original 50°C temperature. But the heat content of a pint of 50°C water is just half the heat content of a quart of 50°C water."

It is incredible that writers and teachers are still referring to the "heat in a body." When water flows, the hydraulic engineer refers to a current or, more specifically, a "water current." Similarly, when electric charge flows, one refers to an electric current, and when heat flows, one alludes to a heat current. There is, however, a fundamental difference: both water and electricity are *matter*, which (neglecting relativistic effects) is conserved. Heat, however, is *not* conserved. When water and electricity enter a system, they don't disappear. When heat enters a system, it has no existence within the system as "heat." Imagine an ideal gas, in contact with a large furnace, undergoing an isothermal expansion. Heat goes in, but the gas undergoes no change of energy, since the work done by the gas compensates the inflow of heat. Heat continues to go in, but there is not only no heat in the gas, but no energy change as well!

The explanation of the foregoing quotations is that the authors are trying to get along without making use of the concept of "internal energy" and without mentioning "work." It is the opinion of the present author that it would be nice if this simplification could be accomplished, but it is quite impossible. The following attempt to avoid the use of the terms "work" and "internal energy" appears at first sight to be legitimate, but it really is not. The first law is written

$$E_{\text{total}} = E_{\sigma} + E_{\theta} + E_{wt},$$

where E_{σ} is the energy of the system, E_{θ} the energy of the "thermal" surroundings and E_{wt} the potential energy of a suspended object. During a process, the three energies on the right undergo changes, but the principle of the conservation of energy requires that E_{total} remain constant. Hence

$$0 = \Delta E_{\sigma} + \Delta E_{\theta} + \Delta E_{wt}.$$

The first term on the right is interpreted as the internal energy change of the system, without any clue as to how it is to be measured, that is, without an operational definition. The identification of ΔE_{θ} with the heat Q and of ΔE_{wt} with the work W are equally unjustified, because E_{θ} is presumably the internal energy of an external system and is therefore a function of its coordinates whose infinitesimal change is an exact differential, whereas dQ is an inexact differential, so that Q depends on the path. Similarly, E_{wt} refers to an external mechanical system and is a function of its mechanical coordinates whose infinitesimal change is an exact differential, in contradistinction to dW which is inexact, so that W depends on the path.

(2) *Using "heat" as a verb.* It is very difficult to avoid the expression "to heat a system." We are accustomed to heating a house, heating a car, heating some water, etc. There are, however, only two possible meanings when "heat" is used as a verb: sometimes it means "add heat to" and at other times "to raise the temperature of"—two expressions that are by no means identical. When we come to the expression used so often by physics and chemistry teachers, namely ". . . if you heat something, it becomes warmer" you get either a redundant statement ("If you raise the temperature of something, it gets warmer"), or an incorrect statement ("If you add heat to something, it gets warmer." What about adding heat to a cake of ice at 0°C?).

(3) *Relying on the undefined concept of thermal energy.* The concept of thermal energy is by all odds the most obscure, the most mysterious, and the most ambiguous term employed by writers of elementary physics and by chemists. Here are only a few examples:

"This brings us to ideas of thermal energy.

In the final chapter, thermal energy is measured

by its heating effect on water."

"The thermal energy of the surrounding is the energy it has in excess of what it would have at absolute zero. . . thermal energy (or heat)."

"Briefly, thermal energy is any energy that is fed into a sample when we raise its temperature, and comes out when the temperature falls. . . The distinction between temperature and thermal energy (or heat) should now be clear."

The main objection to "thermal energy" (aside from the fact that it is undefined), is that at one point it means "heat," and two lines later it means "internal energy." It means whatever pleases the author at the time, and the reader is supposed, somehow, to know which of the two meanings is intended. It is a catch-all expression intended to "simplify" by combining "heat" and internal "energy." Thus to a confirmed user of "thermal energy," there is no flow of "thermal energy" into a gas when it is compressed adiabatically, but a few lines later we learn that the "thermal energy" has increased! During calorimetric experiments with water, "thermal energy" is conserved, but during an isothermal expansion of a gas, it is not. When an isothermal expansion of an ideal gas takes place, the "thermal energy" user admits that there is a transfer of "thermal energy," but no "thermal energy" change, and no temperature change! If the substance is not an ideal gas, an isothermal process involves both a "thermal energy" transfer and a "thermal energy" change, but no temperature change. The contention that "thermal energy" is "any energy that is fed into the sample when we raise its temperature, and comes out when the temperature falls" is wholly inadequate to deal with some of the simplest types of processes studied in elementary physics. As a matter of fact, of the nine processes listed in Table I, only three are amenable to the "thermal energy" treatment. It is much simpler in the long run to prepare the student to accept and to have even a slight feeling for "work," "internal energy change" and "heat," so that he will not be forced to unlearn false ideas in his later work.

I do not advise the study of all the processes listed in Table I during an elementary treatment of thermal physics. The few processes that are unavoidable, such as an adiabatic process, or a phase change, or a cycle should, however, be treated correctly, so that, should another process be brought into discussion, it will be amenable to treatment and not violate what has been said before.

What to Do with Elementary Students

I believe that the rigorous treatment of the first law is well within the reach of the college freshman in America, the A-level student of England, and the gymnasium student on the Continent.

I admit that, at times, I have been guilty of watering down the explanation of internal energy, but students have shown themselves willing to accept the idea of internal energy in terms of molecular kinetic and potential energies. Once they have a feeling for internal energy, they accept work and heat as methods of energy transfer and learn to handle the first law acceptably. There are, of course, several complicated processes to which they could not apply the first law, such as the process of filling an evacuated bottle, or the production of liquid helium by the Simon expansion method. But these can wait for a real course in thermodynamics.

The high-school student is my greatest concern. I have had no experience teaching students who are fourteen, fifteen, or sixteen years old. But I believe that they are willing to accept "internal energy" on a molecular basis. The rigorous definition of internal energy in terms of the negative of the adiabatic work is quite unnecessary. The only problem is to make sure that the child learns to distinguish between the two methods of energy transfer, work and heat. For this purpose I have found the nine diagrams of Fig. 2 helpful for students of all ages. I recommend them heartily.

One more method of treating the first law is worthy of comment. Suppose the old calorimetric definition of heat is introduced, the 15° calorie is carefully defined, and the calorimetric method of measuring heat is described. Suppose then that the student is told that a system undergoes a process in which it proceeds from a given state i to a specified state f , during which both heat Q and work W are measured as carefully as possible.

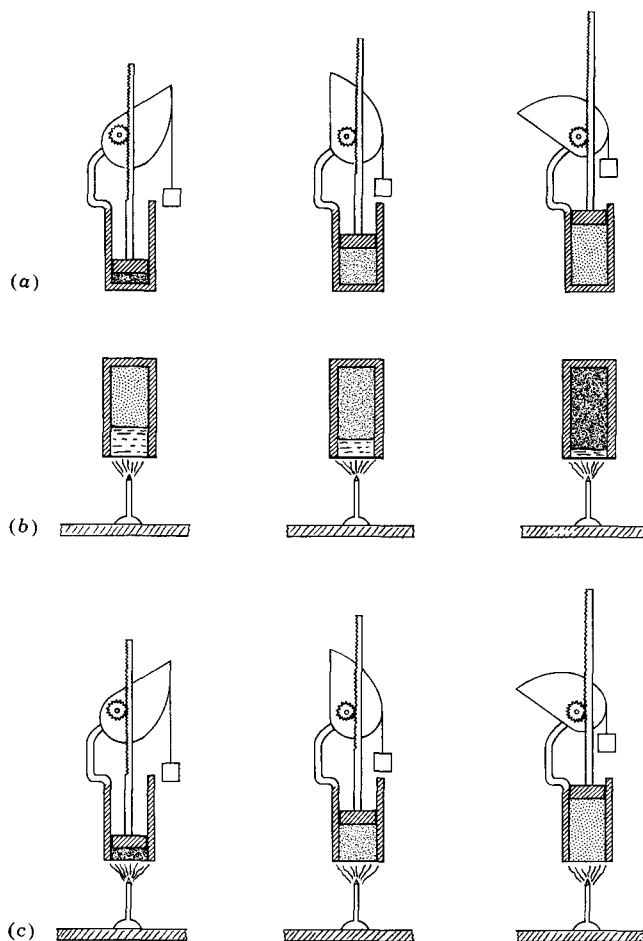


Fig. 2. Distinction between work and heat. (a) adiabatic work; (b) heat flow without work; (c) work and heat.

Table I. Elementary Processes

Process	Heat transfer Q	Internal energy change ΔU	Work W	Temperature change ΔT
Isothermal expansion of an ideal gas	Q	0	W	0
Any isothermal process	Q	ΔU	W	0
Phase change at constant T and P	Q	ΔU	W	0
Any adiabatic process	0	ΔU	W	ΔT
Any cycle	$Q_1 - Q_2$	0	W	0
Any isochoric process	Q	ΔU	0	ΔT
Any isobaric process	Q	ΔU	W	ΔT
Free expansion of an ideal gas	0	0	0	0
Free expansion of a real gas	0	0	0	ΔT

If the student can imagine a number of different ways in which the system may be changed from the same i to the same f in each of which both Q and W are measured, then he may be told that the difference $Q - W$ is found to be the same. Since $Q - W$ is independent of the path from i to f , it must be equal to the difference between two values of a function U , called the internal energy.

I do not object to this treatment because, at the level at which it is used, it is quite impossible to explain adequately its shortcomings. The main objection to this treatment is the acceptance of the calorie as a definite unit of heat, before the existence of an internal energy function is established. By accepting the calorie as a unit, the idea of an internal energy function is really implied, and therefore, the reasoning is circular. If the concept of internal energy is the fundamental stumbling block, it is better to start with it in molecular terms, and to focus the students' attention on Q and W as methods of producing changes in the internal energy. I believe this treatment provides the smoothest transition from elementary thermal physics to legitimate thermodynamics. ■