REVIEWS OF MODERN PHYSICS VOLUME 40, NUMBER 3 JULY 1968

Fundamental Thermodynamics Since Carathéodory

OTTO REDLICH
Inorganic Materials Research Division, Lawrence Radiation Laboratory,
and
Department of Chemical Engineering, University of California, Berkeley, California

The uneasiness prevailing in thermodynamics is the inevitable result of the absence of clearly understood basic concepts. Since thermodynamics, the doctrine of equilibrium, encompasses parts of all branches of the physical sciences, its basic concepts must be explained in ordinary language, free from all references to a specific branch. Moreover, the general applicability of the concepts must be demonstrated.

Carathéodory has introduced new ideas of fundamental importance. But his shortcomings concerning basic concepts have been repeated and intensified by his successors in axiomatics.

A properly constructed basis of thermodynamics elucidates the fundamental distinction of generalized coordinates and generalized forces from each other and from other properties. As an accessorial result one notices a unique quality of the generalized forces (and the temperature): They can be measured only if equilibrium has been established.

The so-called "zeroth law" is neither a law nor a generalized observation. The clarification of its significance contributes to the epistemological understanding of thermodynamics.

CONTENTS

1. SCOPE

Thermodynamics owes its origin to a technical problem, the efficiency of steam engines. It has grown by two different mechanisms: by the logical expansion of limited observations to general laws, and by the extension of these laws to cover new fields. Starting from the compression of a gas, the application of thermodynamics has been extended to surface phenomena, elastic processes, electric and magnetic changes, solutions, phase changes, chemical reactions, and biological and cosmological problems.

Contemplating these steps of progress, widely varying in kind and importance, one wonders if the question of the scope of thermodynamics has been given enough thought. Staking out borders is usually an appallingly sterile activity. In the present case it derives its justification from the astounding generality of thermodynamics and our ensuing obligation of constructing an appropriately general basis. Thermodynamics covers, indeed, our entire knowledge of equilibrium and processes occurring near equilibrium in all fields of physical sciences. In this sense thermodynamics may be called the root of all sciences. The various branches sprout in the problems of kinetics and dynamics. Thus mechanical dynamics or electrodynamics or chemical kinetics are independent branches, except that their concepts must be concordant with thermodynamical concepts in the special case of equilibrium.

The width of scope obliges us to express and discuss fundamental thermodynamics by means of concepts of equally general applicability. To some degree this requirement has been felt by several authors. The present survey is intended to show how it can be satisfied.

Immediately the question arises, *how* one can introduce such concepts and, moreover, how one can *demonstrate* their general applicability. In physical science, a concept is defined by an experimental instruction. For quantitative concepts, the instruction must lead to a Dedekind cut, i.e., a measurement. It is necessary and sufficient for the definition of a property P that the prescribed experimental procedure decides whether the value P_A of P in an object A is greater than, equal to, or smaller than the value P_B of P in an object B. A definition of this kind covers any number of a group of functions transformable into one another by a monotonically increasing transformation. The selection of a particular member of this group as a measure of the property is essentially arbitrary and a matter of convention. (Instead of the temperature T we could choose $\log T$ as a measure of the same property.)

There remains the question of the general applicability of a concept. If we start with a concept borrowed from mechanics, we are not entitled to expect that it will be useful in electric phenomena. But how can we find concepts that are generally applicable and, moreover, reveal themselves as being so? Essentially, Kant has given us the guideline. In a somewhat free interpretation we may express what we have learned from the *Critique of Pure Reason* in this way: In order to find concepts that are necessarily general, we have to search for those concepts that are indispensable for expressing our observations.

Kant's own search definitely was unsuccessful. But his guideline was sound.

In the following the contributions of Carathéodory and later authors to the foundations of thermodynamics are discussed. We then resume the investigation of the basic concepts.

2. CARATHEODORY

2.1. Heat and the First Law

As soon as the investigations of Gibbs, Helmholtz, H. A. Lorentz, Duhem, Nernst, and others had expanded the field of thermodynamics, numerous authors felt the need of setting up a rigorous system, free of contradictions. The idea was to derive the two laws and the essential content of thermodynamics from two principles obtained by generalization of observations. The most beautiful realization of this program was presented by Planck (1897).

Mach's influence in those days was strong enough to cause an emphatic accent on the empirical basis. It was not strong enough to produce a logical investigation of the basic concepts such as Mach himself had undertaken in mechanics.

It was Carathéodory (1909) who felt the need of an entirely new system. He may have been led to this problem by the fact that he had previously developed a very suitable tool for discussing the second law in the theory of Pfaffian expressions. But his contribution to a new formulation of the first law was even more important. His mathematically trained mind was offended by a redundancy in the primary concepts.

The origin of thermodynamics had, indeed, entailed the combined use of thermal and mechanical concepts, both based directly on observations. Carathéodory set out to unify the system by reducing thermal notions to mechanical terms. The notions in question are amount of heat and temperature.

Carathéodory leads to a nonthermal presentation of the first law by way of a direct experimental definition of the idea of an adiabatic wall. He notices that the equilibrium state of an object enclosed in a deformable vessel depends on the outside pressure, but that the state of an object enclosed in a rigid vessel is not subject to any condition dependent on the mechanical properties of the environment. In an entirely analogous way there may or may not exist equilibrium conditions that have nothing to do with the outside pressure. There are walls of such a nature that an enclosed object is subject to no other but mechanical equilibrium conditions; such walls are called *adiabatic*; they are approximately realized by a thermos bottle. But there are other walls, called *diathermic*, such that an enclosed object is subject to one other equilibrium condition, *in addition* to any mechanical conditions.

Now the first law can be formulated: If an object enclosed in an adiabatic vessel changes from an initial state I to a final state F, the work W done upon the object is always the same. It may therefore be used to define a state function, the energy E, such that

$$E_F - E_I = W.$$

This may be illustrated by a simple example. A gas is enclosed in a cylinder equipped with a piston and an electric heating coil. The cylinder is kept in a thermos bottle. If we compress the gas and let it then expand to the original volume without doing any work, we attain a final state in which the gas is warmer. The same state can be attained by introducing electric work into the heating coil. The amount of work required to attain the same final state is the same in both cases.

The amount of heat introduced into an object in a diathermic vessel is now defined as

$$Q = E_F - E_A - W. \tag{2.1}$$

It is zero for an adiabatic change.

Temperature is introduced as a state function such that equality of the temperatures of the object and the environment represents the nonmechanical equilibrium condition for an object in a diathermic vessel.

Thus the thermal concepts are indeed reduced to mechanical ones so that the unity of the system is established.

Epistemologically the great progress results from the elimination of heat as a fundamental concept that is based directly on observation. Whenever such a redundant concept is introduced, it must be later removed; this is done again by appeal to experiment.

In the present problem, Carathéodory eliminated the unnecessary appeal by reducing the principle of the first law to a statement concerning only adiabatic changes. The redundant system covers up the lack of a straightforward, logical presentation by a double recourse to observation.

In the question of heat Carathéodory pioneered the cleanup of thermodynamics. [Carathéodory mentioned that G. B. Bryan (1906) and J. Perrin (1906) had noticed the redundancy before. I have not found any statement or hint of a pertinent nature in these two papers.] But he left a good deal undone.

2.2. Entropy and the Second Law

In the discussion of the second law, Carathéodory again started an important straightening process. Here he eliminated the use of such artificial devices as the Carnot cycle or a periodically operating machine. These devices, historically well justified, never did excel in clarity, directness, and conspicuous generality of argument.

Carathéodory derives the second law from the following empirical principle: In the neighborhood of any state A of an object there are other states B that cannot be reached from A in *adiabatic* processes. For example, the temperature of a galvanic cell in a thermos bottle may be increased by introducing electric work into a heating coil without charging or discharging the cell. But there is no way of adiabatically decreasing the temperature, directly or indirectly. To be sure, a discharge of the cell may be coupled with a temperature decrease but restitution of the original charge leads again to the original (or a higher) temperature, never to a lower temperature at the same state of charge.

The principle that such adiabatically inaccessible states exist is introduced by Carathéodory as a generalization of experimental findings. It leads to the definition of entropy. The first law furnishes a differential expression (Pfaffian expression)

$$đQ = dE - đW \tag{2.2}$$

for the heat introduced if the process is quasistatic, i.e., a sequence of equilibrium states. In this case the work $đW$ introduced can be expressed by properties of the object. (If equilibrium is not established, not all properties of the object have well-defined values and $đW$ can be expressed only by properties of the environment.)

Such a Pfaffian expression as Eq. (2.2) is, in general, except for trivial cases, not integrable. But if the curves defined by

$$đQ = dE - đW = 0 \tag{2.3}$$

do not lead to all points in the neighborhood of the starting point, then (2.3) is integrable. In other words, according to a theorem of Carathéodory for Pfaffians, the condition (2.3) for an adiabatic-quasistatic change is then equivalent to the constancy of a state function S given by

$$dS = đQ/T = (dE - đW)/T = 0. \tag{2.4}$$

The differential dS of S is obtained by multiplying (2.3) by a state function $1/T$. The theorem states the existence of the state functions S and T provided that the family of curves defined by (2.3) does *not* cover *all* neighbor points.

According to (2.4), the entropy S is constant in an adiabatic-quasistatic change. The set of adiabatic-quasistatically accessible states represents the border between adiabatic-irreversibly accessible states and adiabatically inaccessible states. Conventionally irreversible changes are characterized by

$$dS > 0, \tag{2.5}$$

so that the entropy is sufficiently defined by a Dedekind cut. It is measured in diathermic changes by the quasistatically introduced heat as

$$dS = đQ/T. \tag{2.6}$$

Since the definition of T cannot include the value zero, T must be defined either as always positive or always negative. Consideration of thermal interaction identifies T as the temperature.

3. LATER DEVELOPMENT

In the twelve years following Carathéodory's paper his achievements in thermodynamics were hardly noticed. But when Born (1921) had published a simplified presentation combined with an emphatic appreciation, Carathéodory's ideas were widely studied and accepted.

An important modification of the second law has been proposed by Buchdahl (1958). He introduces as a generalized empirical principle the classification of the states of an object. There is a class of states B which are the results of adiabatic changes starting from state A while no adiabatic change leads from B to A (adiabatic-irreversible changes). There is a second class of states C which is not adiabatically accessible from A. The third class of states D is adiabatically accessible from A, and the state A is also adiabatically accessible from any of the states D (adiabatic-reversible changes).

This principle is sufficient for a Dedekind cut and therefore the definition of a state function S which has the same value for A and all states D (conventionally), higher values for the states B, and (accordingly) lower values for the states C.

The gain in clarity and direct emphasis of the essential content of the second law is obvious.

Carathéodory's attempt to present thermodynamics on an axiomatic basis, has been taken up by several authors [Ehrenfest-Afanassjewa (1925), Landsberg (1956, 1961), Falk (1959), Falk and Jung (1959), and Callen (1960)].

Carathéodory's work demonstrates the importance of a thorough mathematical basis in the presentation of thermodynamics. But the reviewer has not been able to find equally significant results in the discussions of his successors. The basic concepts of thermodynamics and their physical meaning have not been given much thought; the introduction of the concepts, following traditional lines, contains errors. (Examples are discussed in Sec. 5.3.) Thus axiomatics in thermodynamics has remained a game for chips that are not interconvertible with real money. Mathematicians sometimes are said to prefer their science *because* it cannot find any practical use. There is no need here to discuss either the factual basis or the moral value of this statement; but it is obvious that this viewpoint is incompatible with the objective of applying axiomatics to thermodynamics.

4. BASIC CONCEPTS

Up to this point no attempt has been made to introduce precise concepts in an orderly fashion. Since the purpose of the preceding sections was a report of the work of Carathéodory and other authors, such an attempt would have been pedantic and even unfair. At this point, however, we have to resume the discussion of the first section so that we are prepared then to realize the achievements and the shortcomings of earlier authors in regard to the basic concepts.

The significance and requirements of an orderly introduction of the basic concepts have been indicated in the first section. The nihilist attitude of jumping into a discussion of terms whose meaning never has been explained obviously does not make sense in physical sciences. It has never been openly advocated, yet it has been silently adopted by all writers in thermodynamics without a single exception. Denbigh (1955) has been more candid than others by frankly passing the buck: "The notion of work is not regarded as being in need of definition in thermodynamics, since it is a concept which is already defined by the primary science of mechanics." Since thermodynamics has a much wider scope than mechanics, a mechanical definition can never do. Zemansky's (1957) explanation, going back to Poincaré, namely, that work can be used to lift a weight, is entirely vague and cannot be translated into operative terms.

The method of introducing basic concepts has been outlined a few years ago (Redlich, 1962). The meaning of the concepts must be explained in ordinary language; no terms of a particular science must be used, but the concepts must be applicable in all physical sciences. The discussion may be illustrated by examples but it must not be based on examples. The inevitability of the introduced concepts must be shown and therefore no concept must be defined by enumeration of particular examples.

The goal of natural science is the description of reproducible events. Thus they are distinguished from history, which describes unique events. The experimental nature of the physical sciences requires a particular distinction of properties that can be varied according to our pleasure. A description using such properties ensures the experimental reproducibility of the conditions of observation.

Our description of the world, the so-called natural laws, is tentative. If an event is not correctly predicted, the underlying "natural law" must be amended. The refutability of all results of natural science by new observations, in particular by experiment, has been clearly pointed out by Popper (1935) as a decisive characteristic.

On this basis a set of basic concepts can be developed. A brief outline follows.

4.1. Object and Isolation

The concepts "object" and "isolation" are indispensable because we cannot describe the whole universe in a single swoop. The description can proceed only piecemeal. The two concepts are coupled: "Object" is anything that can be isolated, and an "isolated" object is one whose properties remain unchanged whatever changes may happen in its environment.

In these concepts as well as in all others, we permit ourselves far-reaching idealization. But legitimate idealizations start from experimental situations. The "isolated object," for instance, is an idealization of a constant-volume calorimeter. A properly constructed concept eliminates the clumsy language that would be required in a direct description of an experiment with all its shortcomings. An idealized calorimeter, i.e., the container of an isolated object, would be described by a whole series of sequences, walls of decreasing thickness at constant rigidity, vacuum jackets of decreasing pressure, supports of decreasing cross section, and so on. The imperfect realization of the experimental conditions does not prevent us from using idealized concepts such as "isolated object" and numerous others.

4.2. Interaction by Contact

After the first step of studying isolated objects, the second is obviously to examine interaction between two otherwise isolated objects. Empirically we notice that interaction exists always whenever two objects touch each other. One condition of isolation is enclosure in a vacuum jacket, such as a thermos bottle. Interaction by contact is also called thermal interaction.

If an object A in contact with an object B becomes warmer, we say $T_A < T_B$ and conversely. The observation of thermal interaction constitutes therefore the basis for a Dedekind cut and thus for the definition of the temperature T.

4.3. The "Zeroth Law" and Nonthermal Interaction

Carathéodory believed that the definition of temperature requires, as an empirically based condition, the statement: "If an object A is in thermal equilibrium with B ($T_A = T_B$), and if A is in equilibrium with C, then B and C are always in equilibrium with each other." This condition has been called the "zeroth law" by Fowler and Guggenheim (1939), a term repeated by numerous authors. It is strange that no author has noticed that an analogous condition should be presupposed for the concepts of the mechanical force, of voltage, of the chemical potential and other quantities.

The "zeroth law" is not the generalization of observations, it is not a necessary condition for the definition of temperature, and it is no law. Its real significance can be illustrated by the following example. We choose an object A which is permeable for neutrons, an object B that absorbs neutrons, and an object C that radiates neutrons. Not knowing anything of these radiation properties, we establish thermal equilibrium between A and B, and between A and C. Then we find that B warms up on contact with C.

Do we conclude that the concept of temperature is meaningless? By no means. We conclude that there is a new, nonthermal mode of interaction and set out to describe isolation and the particular conditions of interaction for this new mode.

A more conventional example would be the choice A = water, B = benzene, C = carbon tetrachloride. The thermal equilibria AB and AC are easily established, but B and C produce a heat of mixing on contact.

The generalization is obvious: Whenever the so-called zeroth law is invalid, we have to search for a new mode of interaction.

Each new mode leads to some particular interaction gadgets which permit us to establish or eliminate interaction between two objects. These may be a mechanical connection for mechanical interaction, or a pair of copper wires and a switch for electric interaction, or a semipermeable membrane for mixing and chemical reactions, and so on.

Instead of a "law" expressing a generalization of observed facts, we have a requirement, a "rule of order" imposed by us on any description of natural events.

Popper's criterion of refutability shows immediately that the "zeroth law" is not an empirical statement: It can never be found to be invalid. If it is taken as a factual statement, one is forced to introduce in its application conditions of isolation (Carathéodory's various "walls") that come as afterthought. As introduced at the start, they are entirely indefinite. Restrictions imposed afterwards in particular cases have sapped the conceptual strength of thermodynamics.

4.4. Interaction Condition: Generalized Coordinates

What does "establishing nonthermal interaction" between two objects mean? Simple examples are easily given. In the interaction between two weights on a balance it means releasing the arresting mechanism; in the uniting of two gases it means opening the stopcock in the connecting tube; in the interaction between a galvanic cell and a capacitor it means closing a switch. But what is the general significance of such an operation in a quantitative description?

The properties of an isolated object are independent of the properties of another isolated object. Accordingly, it is the general characteristics of interaction that a condition

$$F(x', x'') = 0 \tag{4.1}$$

is imposed on a property x' of the first object and a property x'' of the second one. Since we can replace any quantity by a monotonically increasing function, we may always transform the interaction condition (4.1) to

$$x' + x'' = \text{const} \tag{4.2}$$

or

$$dx' + dx'' = 0. \tag{4.3}$$

This is usually, though not always, done.

As long as we consider only one variable x' of an object, the particular choice of the property would be of little concern. We could characterize the state of a gas just as well by its dielectric constant as by its volume. This is no longer the case as soon as we consider more than one mode of interaction. A set of h interaction conditions of the kind (4.3) would be of little descriptive value if each of the h conditions did involve the variables x_1', x_2', $\cdots$, x_h' of the first object and the corresponding variables of some other object. For a rational description of physical events and for the purpose of experimentation, we must restrict the choice of the variables: We select as *generalized coordinates* $x_1, x_2, \cdots, x_h$ a set of independent variables of such kind that, in interaction by a certain mode j and isolation with respect to all other modes, only the coordinate x_j is changed while all others remain constant.

This orthogonality restriction is obviously necessary for an orderly description of each mode of interaction and for maintaining our ability of changing the object to an arbitrary state by interaction with other objects. It is a requirement and we have no guarantee that it can be satisfied. We impose it on physical science, find difficulties in satisfying it, and muddle through somehow. This is the natural course of science.

4.5. Equilibrium: Generalized Forces

The result of interaction between two objects may be: (a) increase of x', and therefore, according to (4.3), decrease of x''; (b) decrease of x' and increase of x''; (c) no change in either x' or x''. The observation of interaction is therefore the basis for a Dedekind cut and for the definition of a new property, the *generalized forces* f' and f'' of the two objects.

The generalized forces are defined by

$$\text{(a)}\quad f'' > f' \quad \text{if} \quad dx' > 0;\quad dx'' < 0, \tag{4.4}$$

$$\text{(b)}\quad f'' < f' \quad \text{if} \quad dx' < 0;\quad dx'' > 0, \tag{4.5}$$

$$\text{(c)}\quad f'' = f' \quad \text{if} \quad dx' < 0;\quad dx'' = 0. \tag{4.6}$$

Case (c) defines equilibrium. It differs from isolation of the two objects in that a small change enforced on the second object may entail a change in the first object.

Thus there is a generalized coordinate (path, volume multiplied by -1, electric charge, surface area, and so on) conjugate to each mode of interaction (mechanical, electric, and so on) and also a generalized force (mechanical force, pressure, voltage, surface tension, and so on). The conventional calibration of all forces, starting from the weight of a piece of platinum-iridium, does not require any further discussion.

For the definition of forces and the application of the "zeroth law" we could repeat precisely what was said regarding temperature. If equilibrium between A and B and equilibrium between A and C does not entail equilibrium between B and C, we have to search for a new mode of interaction.

The measurement of forces furnishes a very characteristic distinction of generalized forces (and the temperature) from all other properties. Forces are measured by comparison, as are all quantities. But in order to determine a force, we must establish equilibrium between two objects: The object on the left pan of a balance must be in equilibrium with the standard weight on the right pan, the pressure between an object and the gauge must be balanced, the voltage of a galvanic cell and of the potentiometer must be equalized, chemical potentials are measured in equilibrium (for instance, in the isopiestic method). The thermometer is used in the same manner.

The definition of work done upon the first object in the mode j

$$W_j' = -\int f_j'' dx_j' \tag{4.7}$$

requires only that the force f_j'' of the environment is well defined. There must be equilibrium between the environment and a gauge, but not necessarily between the environment and the object.

The situation is different for the determination of the entropy. Here the temperature of the object must be defined, but no reversibility is required in the environment.

5. GENERALIZED COORDINATES AND FORCES IN THE DEVELOPMENT OF THERMODYNAMICS

The outline of the preceding section furnishes the background for a review of the basic concepts in the thermodynamic literature. The important concepts are the generalized coordinates and forces.

5.1 Before Carathéodory

In early thermodynamics the only coordinate considered was the volume and accordingly only pressure was introduced as a force. One by one other coordinates and forces were examined. Textbooks sometimes developed the fundamental laws referring only to work against pressure; other modes of interaction were often added as an afterthought. The irrational change of scope was one of the main sources of a widely spread feeling of uncertainty.

Only one of the early authors, Helm (1898), made a serious attempt at discussing the properties that we now call generalized coordinates and forces. He tried to characterize and enumerate them and called them "extensities" and "intensities," respectively. He was not able to introduce these concepts on a firm basis, as has been done in the preceding section.

Helm and the terms used by him were almost completely forgotten early in this century. Undoubtedly they were unknown to Tolman (1917) when he proposed the terms "extensive" and "intensive" and defined them in the manner adopted a little later by Lewis and Randall and generally accepted today. [Apparently neither Tolman nor Lewis had noticed that Planck (1897) had used the terms "external" and "internal" variables in a similar sense.]

In the last twenty years the terms "generalized coordinates" and "forces" came into use, although nobody was able to formulate their meaning. In this situation, obviously by some vague reminiscence of Helm's terms, coordinates were often called extensive and forces intensive. Actually the two pairs of concepts have nothing in common. The simplest example (suggested to the author by Dr. Martin G. Redlich twenty years ago) is a weight in the gravitational field of the earth. The coordinate (the altitude) is intensive, and the force (the weight) is extensive.

5.2. Carathéodory

Through the veil of an austere language, we gain a glance now and then at Carathéodory's personal attitude to his subject. In very crude words, we may perhaps interpret his attitude in the following manner: "What thermodynamics needs is the establishment of logical order, essentially an intellectual cleanup. This is a problem for a mathematician. The fundamental ideas and concepts have been introduced by the physicists long ago and a mathematician need not worry about them."

Accordingly Carathéodory defines the state of a liquid or gaseous phase by the amounts of its constituents, its volume, and its pressure, without giving these terms any thought. He expressly excludes crystalline phases, gravitational and other fields, and electromagnetic and surface forces.

In all these restrictions he refers to the example of Gibbs (1876). But his intention was entirely different. Gibbs did not wish to derive and discuss the general laws but his intention was to *apply* them to certain problems that he precisely circumscribed at the start and then, one by one, treated exhaustively. Consequently, Gibbs talked of energy and entropy as well-known quantities *in the very first sentence* of his paper. Carathéodory wished to develop a general system of thermodynamics. In this endeavor no restrictions at all can be justified. At least, he mentions some of the restrictions in the beginning while other authors have been silent about them.

It may be objected that Carathéodory's terms should not be taken narrowly. "Volumes," for instance, should be taken as a symbolic term covering what in the preceding section was called "generalized coordinates." But such an interpretation is not compatible with Carathéodory's clear language and with his reference to Gibbs, who introduced the same restrictions where they were perfectly proper. Moreover, one would hardly extend the concept of a "wall" or "membrane" to include a pair of copper wires and a switch for the interaction between a galvanic cell and a capacitor.

In addition to the initial restrictions concerning the independent variables, more assumptions turn up in the course of Carathéodory's discussion. Neither the introduction of equilibrium conditions nor the assumptions concerning the transformation of variables made dependent by these conditions are transparent. The further restriction to what Carathéodory calls "simple systems" implies assumptions expressed as definitions; neither their significance nor the need for them is immediately clear (cf. Falk and Jung, 1959). All but one of the independent variables of a "simple system" are assumed to depend only on the phase volumes.

Concentrating on the relations between the basic concepts, the axioms, and the final conclusions, Carathéodory took the basic concepts for granted.

The solution of the mathematical problems involved is necessary but not sufficient for the development of a physical science. The most excellent axiomatics is still not thermodynamics.

5.3. After Carathéodory

In general, Carathéodory's successors have accentuated his shortcomings. Born did not see any problem in the generalization of volumes. Consequently, the body of his discussion is restricted to three independent variables (the simplest nontrivial case for the discussion of Pfaffians).

Neither he nor any of the axiomatists had qualms regarding the distinction of generalized coordinates from other variables or concerning the distinction of coordinates and forces. Following Carathéodory, one accepted the definition of the state of a phase (of constant composition) by volume and pressure as an empirical fact. When the temperature was introduced, the existence of an equation of state was again presented as an empirical fact. This arbitrary procedure of introducing and eliminating undistinguished (in no manner characterized) variables by repeated appeal to experience should be compared with the introduction of one independent variable for each mode of interaction. A systematic deduction, meaningful in every single step, does not need to replace logical development by arbitrary appeal to observation. But in order to build up a clear system one must distinguish between different kinds of variables, as has been done in Sec. 4.

It was Ehrenfest (1911) who felt that something was fundamentally wrong regarding coordinates and forces: "Eine mich völlig befriedigende Definition dieser Begriffe habe ich weder in der Literatur finden können, noch auch selber zuwege gebracht." And in his concluding remarks Ehrenfest mentions that the distinction between coordinates and forces may need an axiomatic investigation "something of the kind presented recently by C. Carathéodory for other concepts of thermodynamics."

Twenty three years later these remarks led to a discussion [Planck (1934, 1935); Ehrenfest–Afanassjewa and de Haas–Lorentz (1935)] that did not clarify the issue and is forgotten today—strange facts in view of the eminence of the participants. The simple distinction of generalized forces, namely, to be measured only in equilibrium (Sec. 4.5), pointed out only much later (Redlich, 1962), could have immediately resolved the discussion.

It is amazing to see that Carathéodory and (even more so) all later axiomatists take infinite pains in the minute examination of a thousand details and are in no way concerned with the meaning of such terms as work, generalized coordinates, and forces.

Among an abundance of new terms, Falk and Jung (1959) introduce the names "metric variables" and "contact variables". The meaning is not easily discovered, but they use the two terms in the place of generalized coordinates and forces, respectively. Then we find these concepts casually identified with "extensive" and "intensive" variables (p. 120). At first, we conclude that this is just the frequent error discussed in Sec. 5.1. But the confusion goes further. On p. 131 we find: "The connection between a metric variable and a conjugate interaction with a conservation law (as discussed above in the example of energy) is of a *general* nature. *The variables conventionally called extensive are quantities of this kind.*" (Italics in the original.) It was mentioned (Redlich, 1962) that a relation looking like a conservation law results from the conventional form (4.3) of the interaction condition *if* the generalized coordinate is extensive. It was also pointed out that these "conservation laws" are unessential. Indeed, relation (4.3) is valid also for intensive coordinates. Moreover, relations of the form (4.3) are a consequence of conveniently chosen coordinates; any monotonically increasing function of the coordinate could again serve as a coordinate, though much less conveniently; the interaction condition (4.3) would appear in a more complicated form but the essential content would be the same. The mix-up of the (basic) distinction of generalized coordinates and forces with the (convenient but unessential) distinction of extensive and intensive properties leads to an entanglement that, in the end, can hardly be resolved.

The same mix-up permeates many of the recent books on thermodynamics. Landsberg (1961) defines: "Any thermodynamic function f which can be expressed in terms of a complete set of independent thermodynamic variables $X_1, X_2 \cdots$, such that

$$f(aX_1, aX_2, \cdots) = a\cdot f(X_1, X_2\cdots) \qquad (5.1)$$

is called an extensive variable." A weight W at the height h in the gravitational field of the earth is, according to this definition, *not* an extensive quantity since obviously

$$W(ah) \neq a\cdot W(h).$$

Before, Landsberg (1956) had casually used (pp.

373, 374, 379, 380) the term "external parameters" for generalized coordinates. The name "external variables" had been used by Planck (1897) a long time ago for extensive quantities. A special warning would have been indicated. But the new term is in no way explained and no reader could understand that the "external parameters" are a class of variables with very special qualities.

It is hardly necessary to point out that textbooks and papers less carefully written than those mentioned contain similar errors quite frequently. The subterranean uneasiness, admitted by most students as well as teachers of thermodynamics, is due to the prevalent confusion in basic concepts.

6. CONCLUSION

The great achievements of Carathéodory have created a permanent contribution of fundamental importance. The definition of heat based on the first law and the principle of inaccessibility remain essential parts of any system of thermodynamics. A significant step in the development of the second law has been made by Buchdahl by the direct formulation of the basic principle.

Two shortcomings in Carathéodory's system have influenced the later development. Misguided by Gibbs's example in an entirely different problem, he unduly restricted the scope of his discussion from the start. Relying on the physicists' previous work, he took the fundamental concepts for granted though they had never been properly analyzed.

Later authors have never eliminated these shortcomings. Moreover, their work has been seriously impaired by a confusion in nomenclature: The terms "extensities" and "intensities," coined by Helm (1898) for today's "generalized coordinates and forces," have been mixed up with "extensive" and "intensive" properties (Tolman, 1917).

It may be entirely natural that the deep dissatisfaction with the state of thermodynamics has resulted in the modern tendency towards axiomatics. After all, everybody would expect clarification and rigor from mathematization. That these efforts have not brought about the expected result is significant in itself. It is true that the lavish introduction of innumerable new terms, always with a glance at their utility in the derivation rather than at their intrinsic meaning, is a great obstacle to the acceptance and application of axiomatics. But the lack of the desired clarification is undoubtedly the principal cause of our disappointment.

The way to rebuild thermodynamics starts from a discussion of the fundamental concepts. Such concepts cannot be taken over from any particular branch of the physical sciences. Their general applicability can be ensured with the aid of an idea going back to Kant: Those concepts are general that are indispensable in the description of all observations. A system of such concepts is briefly outlined in Sec. 4. These concepts are idealizations; they cannot be exactly realized. We use them because there is no other way for science. They are not based on observation but represent the "rules of order" that we impose on the process of describing the world. They are indispensable for an efficient description.

ACKNOWLEDGMENTS

The author is profoundly obliged to Dr. F. H. Stross for many helpful and clarifying discussions during the last twenty years. This work was done under the auspices of the U. S. Atomic Energy Commission.

BIBLIOGRAPHY

Born, M., 1921, Physik, Z. **22**, 218, 249, 282.
Buchdahl, H. A., 1958, Z. Physik **152**, 425.
——, 1955, Am. J. Phys. **23**, 65.
——, 1954, Am. J. Phys. **22**, 182.
——, 1949, Am. J. Phys. **17**, 41, 44, 212.
Bryan, G. B., 1906, in *Encyclopedia of Mathematical Sciences* (B. G. Teubner, Leipzig), Vol. V, Pt. 3, p. 81.
Callen, H. B., 1960, *Thermodynamics* (John Wiley & Sons, Inc., New York).
Carathéodory, C., 1909, Math. Ann. **67**, 355.
Denbigh, K., 1955, *The Principle of Chemical Equilibrium* (Cambridge University Press, London), p. 14.
Ehrenfest, P., 1911, Z. Physik. Chem. **77**, 237.
Ehrenfest-Afanassjewa, T., 1925, Z. Physik **33**, 933.
—— and G. L. deHaas-Lorentz, 1935, Physica **2**, 743.
Falk, G. 1959, Naturwiss. **16**, 481.
—— and H. Jung, 1959, in *Encyclopedia of Physics* (Springer-Verlag, Berlin), Vol. III, Pt. 2.
Fowler, R. H., and E. A. Guggenheim, 1939, *Statistical Thermodynamics* (Cambridge University Press, New York).
Gibbs, J. W., 1876, Trans. Connecticut Acad. **3**, 108.
Helm, G., 1898, *Die Energetik* (Veit & Co., Leipzig).
Landsberg, P. T., 1956, Rev. Mod. Phys. **28**, 363.
—— 1961, *Thermodynamics* (Interscience Publishers, Inc., New York).
Lewis, G. N., and M. Randall, 1922, *Thermodynamics* (McGraw-Hill Book Co., New York), Chap. II.
Perrin, J., 1906, Soc. Franc. Phil. Bull. **6**, 81.
Planck, M., 1897, *Vorlesungen über Thermodynamik* (Veit & Co. Leipzig).
——, 1934, Ann. Phys. (5) **19**, 649.
——, 1935, Physica **2**, 1029.
Popper, K. R., 1935, *Die Logik der Forschung* (Julius Springer-Verlag, Vienna).
——, 1963, *Conjectures and Refutations* (Basic Books, New York).
Redlich, O., 1962, J. Phys. Chem. **66**, 585.
Tolman, R. C. 1917, Phys. Rev. **9**, 237.
Zemansky, M. W., 1957, *Heat and Thermodynamics* (McGraw-Hill Book Co., New York).