Removing Heat and Conceptual Loops from the Definition of Entropy

E. Zanchini$^1$ and G. P. Beretta$^{2*}$

$^1$Università di Bologna, Dipartimento di Ingegneria Energetica, Nucleare e del Controllo Ambientale
viale Risorgimento 2, 40136 Bologna, Italy

$^2$Università di Brescia, Dipartimento di Ingegneria Meccanica e Industriale
via Branze 38, 25123 Brescia, Italy

E-mail: $^1$enzo.zanchini@unibo.it; $^2$beretta@ing.unibs.it

Abstract

A rigorous and general logical scheme is presented, which provides an operative non-statistical definition of entropy valid also in the nonequilibrium domain and free of the usual conceptual loops and unnecessary assumptions that restrict the traditional definition of entropy to the equilibrium domain. The scheme is based on carefully worded operative definitions for all the fundamental concepts employed, including those of system, state of a system, isolated system, separable system, systems uncorrelated from each other, environment of a system, process and reversible process. The treatment considers also systems with movable internal walls and/or semipermeable walls, with chemical reactions and and/or external force fields, and with small numbers of particles. The definition of entropy involves neither the concept of heat nor that of quasistatic process; it applies to both equilibrium and nonequilibrium states. Simple and rigorous proofs of the additivity of entropy and of the principle of entropy nondecrease complete the logical framework.

Keywords: Definition of entropy, Second Law, foundations of Thermodynamics.

1. Introduction

As is well known, classical thermodynamics was developed during the 19th century, due to the pioneering contributions by Carnot, Mayer, Joule, Kelvin, Clausius, Maxwell and Gibbs. In 1897, Planck (Planck 1927) stated the Second Law in the form that is still used in most textbooks and is called Kelvin-Planck’s statement of the Second Law: it is impossible to construct an engine which, working in a cycle, produces no effect except the raising of a weight and the cooling of a heat reservoir. In 1908, Poincaré (Poincaré, 1908) presented a complete structure of classical thermodynamics. The basic approach of Poincaré thermodynamics is still used in several university textbooks, with very small changes. In this approach, the First Law is stated as follows: in a cycle, the work done by a system is proportional to the heat received by the system. In symbols, for a cycle

$$Q = JW,$$

(1)

where $J$ is a universal constant which depends only on the system of units. From Eq. (1) it is easily deduced that, in a process of a system $A$ from the initial state $A_1$ to the final state $A_2$, the quantity $JQ - W$ depends only on the states $A_1$ and $A_2$. Then, one defines the energy difference between $A_2$ and $A_1$ as the value of $JQ - W$ for $A$ in the process, i.e.,

$$E_2^A - E_1^A = (JQ - W)_1^2.$$

(2)

Clearly, Eq. (2) is vitiated by a logical circularity, because it is impossible to define heat without a previous definition of energy. The circularity of Eq. (2) was understood and resolved in 1909 by Carathéodory (Carathéodory 1909), who defined an adiabatic process without employing the concept of heat and stated the First Law as follows: the work performed by a system in any adiabatic process depends only on the end states of the system. So, the first conceptual loop in classical thermodynamics, namely the use of the concept of heat in the definition of energy, was removed. Carathéodory proposed also a new statement of the Second Law, (in terms of adiabatic accessibility) which, however, is now used only in a few axiomatic treatments.

In 1937 Fermi (Fermi 1937) presented a well-known treatment of classical thermodynamics. In this treatment, Carathéodory’s statement of the First Law is employed and rigorous theorems are used to define the thermodynamic temperature of a heat source and the entropy of a system. However, some unsatisfactory aspects still remain: the unnecessary concept of empirical temperature is used; the concepts of heat and of heat source are not defined rigorously; a reversible process is defined as a sequence of stable equilibrium states, i.e., as a quasistatic process. Moreover, an incompleteness is present in the definition of the thermodynamic temperature of a heat source. Indeed, the definition is based on a theorem, in which Fermi considers a reversible cyclic engine which absorbs a quantity of heat $Q_2$ from a source at (empirical) temperature $T_2$ and supplies a quantity of heat $Q_1$ to a source at (empirical) temperature $T_1$. He states that if the engine performs $n$ cycles, the quantity of heat subtracted from the first source is $nQ_2$ and the quantity of heat supplied to the second source is $nQ_1$. Thus, Fermi assumes implicitly that the quantity of heat exchanged in a cycle between a source and a reversible cyclic engine is independent of the initial state of the source. This incompleteness in the deductive scheme of thermodynamics is resolved only in the treatment presented here.

A few decades after Fermi’s contribution, two schools of thermodynamics produced relevant further developments. On one hand, the Prigogine school (see, e.g., Prigogine,

On the other hand, the Keenan school deepened the conceptual foundations of thermodynamics and strengthened the bridge between quantum mechanics and thermodynamics. Some improvements of the logical foundations of thermodynamics due to the Keenan school are as follows. Hatsopoulos and Keenan (Hatsopoulos and Keenan, 1965) analyzed deeply the meaning of the Kelvin-Planck statement of the Second Law. They pointed out that, with the term reservoir, Planck did not mean a system in either metastable or unstable equilibrium, but a system in stable equilibrium; otherwise, the statement of the Second Law would be false. However, when a stable equilibrium state is defined rigorously, the Kelvin-Planck statement becomes a corollary of the definition. They called stable equilibrium a state from which a finite change of state of the system cannot occur without a corresponding finite permanent change of the state of the environment; then, they proved a generalized form of the Kelvin-Planck statement of the Second Law as a consequence of the definitions of stable equilibrium state and of normal system. Thus, they removed the second conceptual loop in classical thermodynamics, i.e., the circularity in the Kelvin-Planck statement.

Hatsopoulos and Keenan stated the Second Law as follows: A system having specified allowed states and an upper bound in volume can reach from any given state a stable state and leave no net effect on the environment (Gyftopoulos and Beretta, 2005, p.34, p.373). They also removed from the logical framework of thermodynamics the use of the unnecessary concept of empirical temperature. Indeed, they showed that thermodynamic temperature can be defined directly, without a previous definition of empirical temperature. They also tried to remove the concept of heat from the definition of entropy. Indeed, they presented the definition of entropy in two ways: the first through the concept of heat (which they defined rigorously); the second without the concept of heat. The second definition, however, was incomplete, because according to it the entropy difference between two states of a system could be measured only by means of a standard thermal reservoir, chosen once and for all.

Gyftopoulos and Beretta (Gyftopoulos and Beretta, 2005) completed the definition of entropy outlined by Hatsopoulos and Keenan. They presented a treatment of thermodynamics in which the definition of entropy is not based on the concepts of heat and of quasistatic process, so that the definition applies, potentially, also to local nonequilibrium states. They also broadened and made more rigorous the set of the basic definitions on which the theory of thermodynamics is based.

The increasing interest in nonequilibrium thermodynamics, as well as the recent scientific revival of thermodynamics in quantum theory (quantum heat engines (Scully 2001 and 2002), quantum Maxwell demons (see, e.g., Lloyd, 1989, 1997 and Giovannetti et al., 2003), quantum erasers (Scully et al., 1982, Kim et al., 2000), etc.) and the recent quest for quantum mechanical explanations of irreversibility (see, e.g., Goldstein et al., 2006, Bennett 2008, Lloyd 2008, Maccone 2009), suggest the need for further improvements of the treatment presented in (Gyftopoulos and Beretta, 2005), in order to obtain a rigorous and general treatment of the foundations of thermodynamics which, by the simplest possible conceptual scheme, extends the definition of entropy to the nonequilibrium domains and, being compatible with the quantum formalism, is suitable for unambiguous fundamental discussions on Second Law implications, even in the framework of quantum theory.

In the present paper, Ref. (Gyftopoulos and Beretta, 2005) is assumed as a starting point and two further objectives are pursued. The basic definitions of system, state, isolated system, separable system, environment of a system and processes are further deepened, by developing the logical scheme outlined in (Zanchini 1988 and 1992). The operative and general definitions of these concepts as presented here are valid also in the presence of internal semipermeable walls and reaction mechanisms. Moreover, the treatment in (Gyftopoulos and Beretta, 2005) is, on one hand, simplified by identifying the minimal set of definitions, assumptions and theorems which yield the definition of entropy and the principle of entropy nondecrease in a more direct way. On the other hand, the definition of a reversible process is given with reference to the concept of scenario; the latter is the largest isolated system whose subsystems are available for interaction, for the class of processes under consideration. In this way, the operativity of the definition is improved and the treatment becomes also more explicitly compatible with old (see, e.g., Beretta et al., 1984, Hatsopoulos and Beretta, 2008) and recent (see, e.g., Goldstein et al., 2006, Bennett 2008, Lloyd 2008, Maccone 2009) interpretations of entropy and irreversibility in the quantum theoretical framework.

Finally, we emphasize that the fast growing field of nonequilibrium thermodynamics (see, for instance, Ottinger and Grmela, 1997, Müller and Ruggeri, 1998, Jou, Casas Vázquez, and Lebon, 2001, Vilar and Rubi, 2001, Kjelstrup and Bedeaux, 2008) would rest on shaky grounds without an operative definition of entropy valid also for nonequilibrium states. Indeed, research advances in nonequilibrium thermodynamics span from theory to applications in a variety of diverse fields, and seem to substantiate from many perspectives the validity of a general principle of maximum entropy production (for a recent review, see Martyushev et al., 2006) wherein a clear understanding of the definition of entropy for nonequilibrium states appears to be an obvious prerequisite.

2. Basic Definitions

Constituents, amounts of constituents. We call constituents the material particles chosen to describe the matter contained in any region of space \( R \) at a given instant of time \( t \). Examples of constituents are: atoms, molecules, ions, protons, neutrons, electrons. Constituents may combine and/or transform into other constituents according to a set of model-specific reaction mechanisms. We call amount of constituent \( i \) in any region of space \( R \) at a given instant of time \( t \), the number of particles of constituent \( i \) contained in \( R \) at time \( t \).

Region of space which contains particles of the \( i \)-th constituent. We will call region of space which contains particles of the \( i \)-th constituent a connected region \( R_i \) of physical space (the three-dimensional Euclidean space) in which particles of the \( i \)-th constituent are contained. The boundary surface of \( R_i \) may be a patchwork of walls, i.e., surfaces impermeable to particles of the \( i \)-th constituent, and ideal surfaces...
Constituents and constraints:

![Diagram of matter with walls](image)

Collection of matter:

![Diagram of collection of matter with walls](image)

**Figure 1.** Collection of matter with two nonreactive constituents and two internal semipermeable membranes: the overlapping regions of space $R_1$ and $R_2$ are split, for clarity.

(permeable to particles of the $i$-th constituent). The geometry and the permeability features of the boundary surface of $R_i$ (walls, ideal surfaces) can vary in time, as well as the number of particles contained in $R_i$.

**Collection of matter.** We call collection of matter, denoted by $C_i$, a set of particles of one or more constituents which is described by specifying the allowed reaction mechanisms between different constituents and, at any instant of time $t$, the set of $r$ connected regions of space, $R^A = R^1_i, ..., R^r_i$, each of which contains $n^i_r$ particles of a single kind of constituent. The regions of space $R^A$ can vary in time and overlap. Two regions of space may contain the same kind of constituent provided that they do not overlap. Thus, the $i$-th constituent could be identical with the $j$-th constituent, provided that $R^1_i$ and $R^2_j$ are disjoint.

**Comment.** This method allows a simple general description of the presence of internal walls and/or internal semipermeable membranes, i.e., surfaces which can be crossed only by some kinds of constituents and not others. An example of the method is illustrated in Figure 1: a collection of matter $C_A$ with constituents $O_2$ and $N_2$, with a movable external wall and with two movable internal membranes, permeable to $O_2$ and to $N_2$ respectively, is represented by two overlapping regions of space, $R_1^1$ and $R_2^1$, each bounded by a movable wall: $R_1^1$ contains $O_2$, while $R_2^1$ contains $N_2$. In the simplest case of a collection of matter without internal partitions, the regions of space $R^1$ coincide at every instant of time.

**Composition.** We call composition of a collection of matter $C_A$, at an instant of time $t$, the vector $n^A$ with $r$ components which specifies the number of particles contained at time $t$ in each region of space $R^A$ of $C_A$.

**Compatible compositions, set of compatible compositions.** We say that two compositions, $n^{A_1}$ and $n^{A_2}$ of a given collection of matter $C_A$ are compatible if the change between $n^{A_1}$ and $n^{A_2}$ or vice versa can take place as a consequence of the allowed reaction mechanisms without matter exchange. We will call set of compatible compositions for a collection of matter $C_A$ the set of all the compositions of $C_A$ which are compatible with a given one, $n^{A_0}$. We will denote a set of compatible compositions by the symbol $(n^{A_0}, \mathbf{v}^A)$, where $\mathbf{v}^A$ is the matrix of the stoichiometric coefficients.

**External force field.** Let us denote by $F$ a force field given by the superposition of the gravitational field $G$, the electric field $E$ and the magnetic field $H$. Let us denote by $\Sigma^A_i$ the union of the regions of space $R^i_A$, in which the constituents of $C^A_i$ are contained, at an instant of time $t$, which will also be called region of space occupied by $C^A_i$ at time $t$. Let us denote by $\Sigma^A_i$ the union of all the regions of space $\Sigma^A_i$, i.e., the whole region of space spanned by the matter and the walls of $C^A_i$, during the time evolution of $C^A_i$.

We call external force field for $C^A_i$ at time $t$, denoted by $F^A_i$, the spatial distribution of $F$ which is measured at time $t$ in $\Sigma^A_i$ if all the constituents and the walls of $C^A_i$ are removed and placed far away from $\Sigma^A_i$. We call external force field for $C^A_i$, denoted by $F^A_i$, the spatial and time distribution of $F$ which is measured in $\Sigma^A_i$ if all the constituents and the walls of $C^A_i$ are removed and placed far away from $\Sigma^A_i$.

**Comment.** We will assume, in the following, that every stationary (i.e., time-independent) external force field is also conservative (i.e., the work performed by the external force field on any particle depends only on the end positions of the particle); indeed, as far as we know, this property holds for every superposition of a stationary gravitational, a stationary electric and a stationary magnetic field in empty space.

**System, properties of a system.** We will call system $A$ a collection of matter $C^A_i$ defined by the initial composition $n^{A_0}$, the stoichiometric coefficients $\mathbf{v}^A$ of the allowed reaction mechanisms, and the possibly time-dependent specification, over the entire time interval of interest, of:

- the geometrical variables and the nature of the boundary surfaces that define the regions of space $R^A_i$;
- the rates $n^{A_0}_i$ at which particles are transferred in or out of the regions of space, and
- the external force field distribution $F^A_i$ for $C^A_i$.

provided that the following conditions apply:

1. an ensemble of identically prepared replicas of $C^A_i$ can be obtained at any instant of time $t$, according to a specified set of instructions or preparation scheme;
2. a set of measurement procedures, $P^A_1, ..., P^A_r$, exists, such that each $P^A_i$ is applied on replicas of $C^A_i$ at any given instant of time $t$, the arithmetic mean $\langle P^A_i \rangle t$ of the numerical outcomes of repeated applications of $P^A_i$ is a value which is the same for every subsample of replicas of $C^A_i$ (the latter condition guarantees the so-called statistical homogeneity of the ensemble); $\langle P^A_i \rangle t$ is called the value of $P^A_i$ for $C^A_i$ at time $t$;
3. the set of measurement procedures, $P^A_1, ..., P^A_r$, is complete in the sense that the set of values $\{\langle P^A_1 \rangle_{t_1}, ..., \langle P^A_r \rangle_{t_r}\}$ allows one to predict the value at time $t$ of any other measurement procedure satisfying condition 2.

Then, each measurement procedure satisfying conditions 2 and 3 is called a property of system $A$, and the set $P^A_1, ..., P^A_r$ a complete set of properties of system $A$.

**State of a system.** Given a system $A$ as just defined, we call state of system $A$ at time $t$, denoted by $A_t$, the set of the values at time $t$ of
isolated system

Σ

the whole region of space

Comment 1. F

Isolated system

tem

our definition of isolated system in a very simple way: a sys-

Comment 2. If energy were already defined, we could state

Σ

AB

Separable closed systems

Consider a composite system $AB$ such that at time $t$ the states $A_i$ and $B_i$ of the two subsystems fully determine the state $(AB)_t$, i.e., the values of all the properties of $AB$ can be determined by local measurements of properties of systems $A$ and $B$. Then we say that systems $A$ and $B$ are uncorrelated from each other at time $t$, and we write the state of $AB$ at time $t$ as $(AB)_t = A_t B_t$.

Environment of a system, scenario. If a system $A$ is a sub-

system of an isolated system $I = AB$, we can choose $AB$ as the isolated system to be studied. Then, we call $B$ the envi-

ronment of $A$, and we call $AB$ the scenario under which $A$ is studied.

Comment. The chosen scenario $AB$ contains as subsystems all and only the systems that are allowed to interact with $A$; all the remaining systems in the universe are considered as not available for interaction.

Process, cycle. We call process for a system $A$ from state $A_1$ to state $A_2$ in the scenario $AB$, denoted by $(AB)_1 \rightarrow (AB)_2$, the time evolution from $(AB)_1$ to $(AB)_2$ of the isolated system $AB$ which defines the scenario. We call cycle for a system $A$ a process whereby the final state $A_2$ coincides with the initial state $A_1$.

Restriction, for brevity. In the following (for brevity) we will consider only closed systems and only states of a closed system $A$ in which $A$ is separable and uncorrelated from its environment. Moreover, for a composite system $AB$, we will consider only states such that the subsystems $A$ and $B$ are separable and uncorrelated from each other.

Reversible process, reverse of a reversible process. A process for $A$ in the scenario $AB$, $(AB)_1 \rightarrow (AB)_2$, is called a reversible process if there exists a process $(AB)_2 \rightarrow (AB)_1$ which restores the initial state of the isolated system $AB$. The process $(AB)_2 \rightarrow (AB)_1$ is called the reverse of process $(AB)_1 \rightarrow (AB)_2$.

Comment. A reversible process need not be slow. In the general framework we are setting up, it is noteworthy that nowhere we state nor we need the concept that a process to be reversible needs to be slow in some sense.

Weight. We call weight a system $M$ always separable and uncorrelated from its environment, such that:

• $M$ is closed, it has a single constituent, with fixed num-

ber of particles and mass $m$, contained in a single region of space whose shape and volume are fixed;

• in any process, the difference between the initial and the final state of $M$ is determined uniquely by the change in the position $z$ of the center of mass of $M$, which can move only along a straight line whose direction, $k$, co-

incides with that of a uniform and stationary external gravitational force field $G = -gk$, where $g$ is a constant gravitational acceleration.

Weight process, work in a weight process. A process of a system $A$ is called a weight process, denoted by $(A_1 \rightarrow A_2)_w$. 

Closed system, open system. A system $A$ is called a closed

system if, at every instant of time $t$, the boundary surface

every region of space $R_A^t$ is a wall. Otherwise, $A$ is called an

open system.

Comment. For a closed system, in each region of space $R_A^t$, the number of particles of the $i$-th constituent can change only as a consequence of allowed reaction mechanisms.

Composite system, subsystems. If systems $A$ and $B$, de-

fined in the same time interval, are such that no region of space $R_A^t$ overlaps with any region of space $R_B^t$, we will say that the system $C$ whose regions of space of are $R_C^t = R_A^t \cup R_B^t$, is the composite of systems $A$ and $B$, and that $A$ and $B$ are subsystems of $C$. Then, we write $C = AB$ and denote its state at time $t$ by $C_t = (AB)_t$.

Isolated system. We say that a closed system $I$ is an isolated system in the stationary external force field $F_I^t$, or simply an isolated system, if during the whole time evolution of $I$: (a) only the particles of $I$ are present in $\Sigma^I$, and (b) the external force field $F_I^t$ is stationary, i.e., time independent.

Comment 1. In simpler words, a system $I$ is isolated if, at every instant of time: no other material particle is present in the whole region of space $\Sigma^I$ which will be crossed by system $I$ during its time evolution; if system $I$ is removed, only a stationary (vanishing or non-vanishing) force field is present in $\Sigma^I$.

Comment 2. If energy were already defined, we could state

our definition of isolated system in a very simple way: a system is isolated if neither mass nor energy crosses its boundary. However, since the definition of energy requires a previous definition of isolated system, we must provide here an operative definition of isolated system which does not employ the concept of energy. Our method is as follows. Consider a system $I$ which, during its whole time evolution, is surrounded by a region of space which contains no material particle, and suppose that only a stationary and conservative force field (vanishing or non-vanishing) is present in this region of space. Then, system $I$ cannot receive or give out either mass or energy. In fact, photons are not present outside the system, because photons are an unsteady electromagnetic field; moreover, the total work done by the external force field on the particles of the system does not change the total potential plus kinetic energy of the system (i.e., the system Hamiltonian).

Separable closed systems. Consider a composite system $AB$, with $A$ and $B$ closed subsystems. We say that systems $A$ and $B$ are separable at time $t$, if at that instant:

• the force field external to $A$ coincides (where defined) with the force field external to $AB$, i.e., $F_{cA}^t = F_{cAB}^t$;

• the force field external to $B$ coincides (where defined) with the force field external to $AB$, i.e., $F_{cB}^t = F_{cAB}^t$.
Definition of Energy for a Closed System

A state $A_0$ of a system $A$, with environment $B$, is called an equilibrium state if:

- state $A_0$ does not change with time;
- state $A_0$ can be reproduced while the state of the environment $B$ of a composite system $AB$ has no net changes.

Stable equilibrium state of a closed system. An equilibrium state of a closed system $A$ is called a stable equilibrium state if it cannot be modified in any process such that neither the geometrical configuration of the walls which bound the regions of space $\mathbf{R}^3$ nor the state of the environment $B$ of $A$ have net changes.

Comment. The definition can be paraphrased as follows. Let $A_{re}$ be a stable equilibrium state of $A$ and let $B$ be the environment of $A$. Suppose that a process $(A_{re}B_1) \rightarrow (A_2B_2)$ took place and that $A_2$ is different from $A_{re}$. Then, no matter what kind of interaction between system and environment occurred: either $B_2$ is different from $B_1$, or the position of some wall which bounds the matter of $A$, in state $A_2$, is different from the position it had in state $A_1$, or both.

3. Definition of Energy for a Closed System

First Law. Every pair of states $(A_1, A_2)$ of a system $A$ can be interconnected by means of a weight process for $A$. The works done by the system in any two weight processes between the same initial and final states are identical.

Definition of energy for a closed system. Proof that it is a property. Let $(A_1, A_2)$ be any pair of states of a system $A$. We call energy difference between states $A_2$ and $A_1$ either the work $W_{12}^{A\rightarrow}$ received by $A$ in any weight process from $A_1$ to $A_2$ or the work $W_{21}^{A\rightarrow}$ done by $A$ in any weight process from $A_2$ to $A_1$; in symbols:

$$E_2 - E_1 = W_{12}^{A\rightarrow} \quad \text{or} \quad E_1 - E_2 = W_{21}^{A\rightarrow}. \quad (4)$$

The First Law guarantees that at least one of the weight processes considered in Eq. (4) exists. Moreover, it yields the following consequences:

(a) if both weight processes $(A_1 \rightarrow A_2)_W$ and $(A_2 \rightarrow A_1)_W$ exist, the two forms of Eq. (4) yield the same result ($W_{12}^{A\rightarrow} = W_{21}^{A\rightarrow}$);

(b) the energy difference between states $A_2$ and $A_1$ depends only on the states $A_1$ and $A_2$;

(c) (additivity of energy differences) consider a pair of states $A_1B_1$ and $A_2B_2$ of a composite system $AB$; then

$$E_2^{AB} - E_1^{AB} = E_2^A - E_1^A + E_2^B - E_1^B; \quad (5)$$

d (energy is a property) let $A_0$ be a reference state of a system $A$, to which we assign an arbitrarily chosen value of energy $E_{A_0}^A$; the value of the energy of $A$ in any other state $A_1$ is determined uniquely by either

$$E_1^A = E_0^A + W_{01}^{A\rightarrow} \quad \text{or} \quad E_1^A = E_0^A + W_{10}^{A\rightarrow} \quad (6)$$

depending on whether state $A_1$ can be interconnected to $A_0$ by means of a weight process $(A_0 \rightarrow A_1)_W$ or $(A_1 \rightarrow A_0)_W$, respectively. Of course, both relations apply when the interconnecting weight processes are possible in both directions. Rigorous proofs of these consequences can be found in (Gyftopoulos and Beretta, 2005) and (Zanchini, 1986).

4. Definition of Entropy for a Closed System

Assumption 1: restriction to normal systems. We call normal system any system $A$ that, starting from every state, can be changed to a nonequilibrium state with higher energy by means of a weight process for $A$ in which the regions of space $\mathbf{R}^3$ occupied by the constituents of $A$ have no net changes. From here on, we consider only normal systems.

Comment. In traditional treatments of thermodynamics, Assumption 1 is not stated explicitly, but it is used, for example when one states that any amount of work can be transferred to a thermal reservoir by a stirrer.

Theorem 1. Impossibility of a PMM2. If a normal system $A$ is in a stable equilibrium state, it is impossible to lower its energy by means of a weight process for $A$ in which the regions of space $\mathbf{R}^3$ occupied by the constituents of $A$ have no net change.

Proof. (See sketch in Figure 3) Suppose that, starting from a stable equilibrium state $A_{re}$ of $A$, by means of a weight process $\Pi_1$ with positive work $W^{A\rightarrow} = W > 0$, the energy of $A$ is lowered and the regions of space $\mathbf{R}^3$ occupied by the constituents of $A$ have no net change. On account of Assumption 1, it would be possible to perform a weight process $\Pi_2$ for $A$ in which the regions of space $\mathbf{R}^3$ occupied by the constituents of $A$ have no net change, the weight $M$ is restored to its initial state so that the positive amount of energy $W^{A\rightarrow} = W > 0$ is supplied back to $A$, and the final state of $A$ is a nonequilibrium state, namely, a state clearly different from $A_{re}$. Thus, the zero-work sequence of weight processes $(\Pi_1, \Pi_2)$ would violate the definition of stable equilibrium state.
Second Law. Among all the states of a system $A$ such that the constituents of $A$ are contained in a given set of regions of space $\mathbf{R}^1$, there is a stable equilibrium state for every value of the energy $E^A$.

**Lemma 1. Uniqueness of the stable equilibrium state.** There can be no pair of different stable equilibrium states of a closed system $A$ with identical regions of space $\mathbf{R}^1$ and the same value of the energy $E^A$.

**Proof.** Since $A$ is closed and in any stable equilibrium state it is separable and uncorrelated from its environment, if two such states existed, by the First Law and the definition of energy they could be interconnected by means of a zero-work weight process. So, at least one of them could be changed to a different state with no external effect, and hence would not satisfy the definition of stable equilibrium state.

**Comment.** Recall that for a closed system, the composition $n^A$ belongs to the set of compatible compositions ($n^{A1}, v^A$) fixed once and for all by the definition of the system.

**Lemma 2.** Any stable equilibrium state $A_{xe}$ of a system $A$ is accessible via an irreversible zero-work weight process from any other state $A_1$ with the same regions of space $\mathbf{R}^1$ and the same value of the energy $E^A$.

**Proof.** By the First Law and the definition of energy, $A_{xe}$ and $A_1$ can be interconnected by a zero-work weight process for $A$. However, a zero-work weight process from $A_{xe}$ to $A_1$ would violate the definition of stable equilibrium state. Therefore, the process must be in the direction from $A_1$ to $A_{xe}$. The absence of a zero-work weight process in the opposite direction implies that any zero-work weight process from $A_1$ to $A_{xe}$ is irreversible.

**Corollary 1.** Any state in which a closed system $A$ is separable and uncorrelated from its environment can be changed to a unique stable equilibrium state by means of a zero-work weight process for $A$ in which the regions of space $\mathbf{R}^1$ have no net change.

**Proof.** The thesis follows immediately from the Second Law, Lemma 1 and Lemma 2.

**Systems in mutual stable equilibrium.** We say that two systems $A$ and $B$, each in a stable equilibrium state, are in mutual stable equilibrium if the composite system $AB$ is in a stable equilibrium state.

**Thermal reservoir.** We call a closed and always separable system $R$ with a single constituent, contained in a fixed region of space, with a vanishing external force field, with energy values restricted to a finite range in which any pair of identical copies of the reservoir, $R$ and $R^d$, is in mutual stable equilibrium when $R$ and $R^d$ are in stable equilibrium states.

**Comment.** Every single constituent system without internal boundaries and applied external fields, and with a number of particles of the order of one mole — so that the simple system approximation as defined in (Gyftopoulos and Beretta, 2005, p.263) applies — when restricted to a fixed region of space of appropriate volume and to the range of energy values corresponding to the so-called triple-point stable equilibrium states, is a thermal reservoir.

Indeed, for a system contained in a fixed region of space, only when three different phases (such as, solid, liquid and vapor) are present, two stable equilibrium states with different energy values have the same temperature (here not yet defined), and thus fulfill the condition for the mutual stable equilibrium of the system and a copy thereof.

**Reference thermal reservoir.** A thermal reservoir chosen once and for all is called a reference thermal reservoir. To fix ideas, we choose water as the constituent of our reference thermal reservoir, i.e., sufficient amounts of ice, liquid water, and water vapor at triple point conditions.

**Standard weight process.** Given a pair of states $(A_1, A_2)$ of a system $A$ and a thermal reservoir $R$, we call standard weight process for $AR$ from $A_1$ to $A_2$ a weight process for the composite system $AR$ in which the end states of $R$ are stable equilibrium states. We denote by $(A_1 R_1 \rightarrow A_2 R_2)^{sw}$ a standard weight process for $AR$ from $A_1$ to $A_2$ and by $(\Delta E^R)^{sw\_AR}$ the corresponding energy change of the thermal reservoir $R$.

**Assumption 2.** Every pair of states $(A_1, A_2)$ of a system $A$ can be interconnected by a reversible standard weight process for $AR$, where $R$ is an arbitrarily chosen thermal reservoir.

**Comment. Statements of the Second Law.** The combination of Assumption 2 with the statement of the Second Law and Lemma 1 given above, forms our re-statement of the Gyftopoulos-Beretta statement of the Second Law (Gyftopoulos and Beretta, 2005, p. 62-63), which, in turn, is a restatement of that introduced by Hatsopoulos and Keenan (Hatsopoulos and Keenan, 1965, p.34, p.373). The motivation for the separation of the statement proposed in (Gyftopoulos and Beretta, 2005) into three parts is as follows: to extract from the postulate a part which can be proved (Lemma 1); to separate logically independent assumptions, i.e., assumptions such that a violation of the first would not imply a violation of the second, and vice-versa.

In addition to the Kelvin-Planck statement discussed above, the well-known historical statements due to Clausius and to Carathéodory unfold as rigorous theorems in our logical scheme. Proofs can be found in (Gyftopoulos and Beretta, 2005, p.64, p.121, p.133).

**Theorem 2.** For a given system $A$ and a given reservoir $R$, among all the standard weight processes for $AR$ between a given pair of states $(A_1, A_2)$ of $A$, the energy change $(\Delta E^R)^{sw\_AR}$ of the thermal reservoir $R$ has a lower bound which is reached if and only if the process is reversible.

**Proof.** Let $\Pi_{AR}$ denote a standard weight process for $AR$ from $A_1$ to $A_2$, and $\Pi_{AR}^{rev}$ a reversible one; the energy changes of $R$ in processes $\Pi_{AR}$ and $\Pi_{AR}^{rev}$ are, respectively,
Theorem 3. Let $R'$ and $R''$ be any two thermal reservoirs and consider the energy changes, $(\Delta E^R)^{\text{yw}}_{A_{1}A_{2}}$ and $(\Delta E^{R''})^{\text{yw}}_{A_{1}A_{2}}$, respectively, in the reversible standard weight processes $\Pi_{AR'}$ and $\Pi_{AR''}$, with the same initial state $R$ of $R'$ and $R''$, respectively. Then the ratio $(\Delta E^{R''})^{\text{yw}}_{A_{1}A_{2}}/(\Delta E^{R'})^{\text{yw}}_{A_{1}A_{2}}$ is positive; 

a) is zero;

b) depends only on $R'$ and $R''$, i.e., it is independent of (i) the initial stable equilibrium states of $R'$ and $R''$, (ii) the choice of system $A$, and (iii) the choice of states $A_{1}$ and $A_{2}$.

Proof of a). With the help of Figure 5, let us suppose that $(\Delta E^{R''})^{\text{yw}}_{A_{1}A_{2}} < 0$. Then, $(\Delta E^{R'})^{\text{yw}}_{A_{1}A_{2}}$ cannot be zero. In fact, in that case the sequence of processes $(\Pi_{AR'}, -\Pi_{AR''})$, which is a cycle for $A$, would be a weight process for $R'$ in which, starting from the stable equilibrium state $R'_1$, the energy of $R'$ is lowered and the regions of space occupied by the constituents of $R'$ have no net changes, in contrast with Theorem 1. Moreover, $(\Delta E^{R''})^{\text{yw}}_{A_{1}A_{2}}$ cannot be positive. In fact, if it were positive, the work done by $R''$ as a result of the overall weight process $(\Pi_{AR'}, -\Pi_{AR''})$ for $R''$ would be

$$W^{R''} = - (\Delta E^{R''})^{\text{yw}}_{A_{1}A_{2}} + (\Delta E^{R'})^{\text{yw}}_{A_{1}A_{2}},$$

where both terms are positive. On account of Assumption 1 and Corollary 1, after the process $(\Pi_{AR'}, -\Pi_{AR''})$, one could perform a weight process $\Pi_{AR'}$ for $R'$ in which a positive amount of energy equal to $(\Delta E^{R''})^{\text{yw}}_{A_{1}A_{2}}$ is given back to $R''$ and the latter is restored to its initial stable equilibrium state. As a result, the sequence $(\Pi_{AR'}, -\Pi_{AR''})$, $\Pi_{AR''}$ would be a weight process for $R''$ in which, starting from the stable equilibrium state $R'_1$, the energy of $R'$ is lowered and the regions of space occupied by the constituents of $R'$ have no net changes, in contrast with Theorem 1. Therefore, the assumption $(\Delta E^{R''})^{\text{yw}}_{A_{1}A_{2}} < 0$ implies $(\Delta E^{R'})^{\text{yw}}_{A_{1}A_{2}} < 0$.

Proof of b). Given a pair of states $(A_{1}, A_{2})$ of a closed system $A$, consider the reversible standard weight process $\Pi_{AR'} = (A_{1}R'_{1} \rightarrow A_{2}R'_{2})^{\text{yw}}_{A_{1}A_{2}}$ for $R'$, with $R'$ initially in state $R'_{1}$, and the reversible standard weight process $\Pi_{AR''} = (A_{1}R''_{1} \rightarrow A_{2}R''_{2})^{\text{yw}}_{A_{1}A_{2}}$ for $R''$, with $R''$ initially in state $R''_{1}$. Moreover, given a pair of states $(A'_{1}, A'_{2})$ of another closed system $A'$, consider the reversible standard weight process $\Pi_{AR'} = (A'_{1}R'_{1} \rightarrow A'_{2}R'_{2})^{\text{yw}}_{A_{1}A_{2}}$ for $A'R'$, with $R'$ initially in state $R'_{1}$, and the reversible standard weight process $\Pi_{AR''} = (A'_{1}R''_{1} \rightarrow A'_{2}R''_{2})^{\text{yw}}_{A_{1}A_{2}}$ for $A'R''$, with $R''$ initially in state $R''_{1}$. With the help of Figure 6, we will prove that the changes in energy of the reservoirs in these processes obey the relation

$$\left(\Delta E^{R''} \right)^{\text{yw}}_{A_{1}A_{2}} / \left(\Delta E^{R'} \right)^{\text{yw}}_{A_{1}A_{2}} = \left(\Delta E^{R''} \right)^{\text{yw}}_{A'_{1}A'_{2}} / \left(\Delta E^{R'} \right)^{\text{yw}}_{A'_{1}A'_{2}}.$$
standard weight process for AR"; whatever the new states of R' and R" are, again system A is brought from A1 to A2 by a reversible standard weight process for AR' and back to A1 by a reversible standard weight process for AR", until the cycle for A is repeated m times. Similarly, ΠA' is a sequence of weight processes for the composite system A'R"R' whereby starting from the end states of R' and R" reached by sequence ΠA1, system A' is brought from A′1 to A′2 by a reversible standard weight process for A'R", then from A′2 to A′1 by a reversible standard weight process for A'R'; and so on until the cycle for A' is repeated n times. Clearly, the composite sequence (ΠA, ΠA') is a cycle for AA'. Moreover, it is a cycle also for R'. In fact, on account of Theorem 2, the energy change of R' in each process ΠAR' is equal to (ΔE^R')_swrev regardless of its initial state and in each process −ΠA'R is equal to −(ΔE^R')_swrev. Therefore, the energy change of R' in the sequence (ΠA, ΠA') is m (ΔE^R')_swrev + n (ΔE^R')_swrev and equals zero on account of Eq. (9). As a result, after (ΠA, ΠA'), reservoir R' has been restored to its initial state, so that (ΠA, ΠA') is a reversible weight process for R". Again on account of Theorem 2, the overall energy change of R" in the sequence is −m (ΔE^R')_swrev + n (ΔE^R')_swrev. If this quantity were negative, Theorem 1 would be violated. If this quantity were positive, Theorem 1 would also be violated by the reverse of the process, (−ΠA', −ΠA). Therefore, the only possibility is that −m (ΔE^R')_swrev + n (ΔE^R')_swrev = 0, i.e.,

\[
m (ΔE^R')_{A1A2} = n (ΔE^R')_{A1A2}.
\]

Finally, taking the ratio of Eqs. (9) and (10), we obtain Eq. (8) which is our thesis.

**Figure 7. Schematic illustration of the processes used to define the temperature of a thermal reservoir.**

**Temperature of a thermal reservoir.** (Figure 7) Let R be a given thermal reservoir and R' a reference thermal reservoir. Select an arbitrary pair of states (A1, A2) of a system A and consider the energy changes \((ΔE^R')_{A1A2}\) and \((ΔE^R')_{A1A2}\) in two reversible standard weight processes from A1 to A2, one for AR and the other for AR', respectively. We call temperature of R the positive quantity

\[
T_R = T_{R'} \left( \frac{(ΔE^R')_A_{A1A2}}{(ΔE^R')_A_{A1A2}} \right),
\]

where \(T_{R'}\) is a positive constant associated arbitrarily with the reference thermal reservoir R'. Clearly, the temperature \(T_R\) of R is defined only up to the arbitrary multiplicative constant \(T_{R'}\). If for R' we select a thermal reservoir consisting of ice, liquid water, and water vapor at triple-point conditions, and we set \(T_{R'} = 273.16\) K, we obtain the Kelvin temperature scale.

**Corollary 2.** The ratio of the temperatures of two thermal reservoirs, R' and R", is independent of the choice of the reference thermal reservoir and can be measured directly as

\[
\frac{T_{R'}}{T_{R'\prime}} = \left( \frac{(ΔE^R')_A_{A1A2}}{(ΔE^R')_A_{A1A2}} \right),
\]

where \((ΔE^R')_A_{A1A2}\) and \((ΔE^R')_A_{A1A2}\) are the energy changes of R' and R" in two reversible standard weight processes, one for AR' and the other for AR", which interconnect the same pair of states (A1, A2).

**Proof.** Let \((ΔE^R')_A_{A1A2}\) be the energy change of the reference thermal reservoir R' in any reversible standard weight process for AR' which interconnects the same states (A1, A2) of A. From Eq. (11) we have

\[
T_{R'} = T_{R'} \left( \frac{(ΔE^R')_A_{A1A2}}{(ΔE^R')_A_{A1A2}} \right), \quad T_{R''} = T_{R'} \left( \frac{(ΔE^R')_A_{A1A2}}{(ΔE^R')_A_{A1A2}} \right),
\]

so that the ratio \(T_{R'}/T_{R''}\) is given by Eq. (12).

**Corollary 3.** Let (A1, A2) be any pair of states of system A, and let \((ΔE^R')_A_{A1A2}\) be the energy change of a thermal reservoir R with temperature \(T_R\) in any reversible standard weight process for AR from A1 to A2. Then, for the given system A, the ratio \((ΔE^R')_A_{A1A2}/T_R\) depends only on the pair of states (A1, A2), i.e., it is independent of the choice of reservoir R and of its initial stable equilibrium state R1.

**Proof.** Let us consider two reversible standard weight processes from A1 to A2, one for AR' and the other for AR"; where R' is a thermal reservoir with temperature \(T_{R'}\) and R" is a thermal reservoir with temperature \(T_{R''}\). Then, equation (12) yields

\[
\frac{(ΔE^R')_A_{A1A2}}{T_{R'}} = \left( \frac{(ΔE^R')_A_{A1A2}}{T_{R'\prime}} \right).
\]
Definition of (thermodynamic) entropy, proof that it is a property. Let \((A_1, A_2)\) be any pair of states of a system \(A\), and let \(R\) be an arbitrarily chosen thermal reservoir placed in the environment \(B\) of \(A\). We call entropy difference between \(A_2\) and \(A_1\) the quantity

\[
S_2^A - S_1^A = -\frac{(\Delta E_{\text{rev}}^{A})_{A_1A_2}}{T_R} \tag{15}
\]

where \((\Delta E_{\text{rev}}^{A})_{A_1A_2}\) is the energy change of \(R\) in any reversible standard weight process for \(AR\) from \(A_1\) to \(A_2\), and \(T_R\) is the temperature of \(R\). On account of Corollary 3, the right hand side of Eq. (15) is determined uniquely by states \(A_1\) and \(A_2\). Let \(A_0\) be a reference state of \(A\), to which we assign an arbitrarily chosen value \(S_0^A\) of the entropy. Then, the value of the entropy of \(A\) in any other state \(A_1\) of \(A\) is determined uniquely by the equation

\[
S_1^A = S_0^A - \frac{(\Delta E_{\text{rev}}^{A})_{A_0A_1}}{T_R}, \tag{16}
\]

where \((\Delta E_{\text{rev}}^{A})_{A_0A_1}\) is the energy change of \(R\) in any reversible standard weight process for \(AR\) from \(A_0\) to \(A_1\), and \(T_R\) is the temperature of \(R\). Such a process exists for every state \(A_1\), on account of Assumption 2. Therefore, entropy is a property of \(A\), defined for every state \(A_1\) of \(A\).

Comment. In view of the growing revival of interest in the field of nonequilibrium thermodynamics, it is worth emphasizing that one of the most important consequences of the above definition — and of that proposed in 1991 in (Gytoptoulos and Beretta, 2005), improved — is that entropy is well and rigorously defined also for nonequilibrium states.

Theorem 4. Additivity of entropy differences. Consider the pair of states \((C_1 = A_1B_1, C_2 = A_2B_2)\) of the composite system \(C = AB\). Then,

\[
S_{A_1B_2}^C - S_{A_1B_1}^C = S_2^A - S_1^A + S_2^B - S_1^B \tag{17}
\]

Proof. Let us choose a thermal reservoir \(R\), with temperature \(T_R\), and consider the sequence \((\Pi_{AR}, \Pi_{BR})\) where \(\Pi_{AR}\) is a reversible standard weight process for \(AR\) from \(A_1\) to \(A_2\), while \(\Pi_{BR}\) is a reversible standard weight process for \(BR\) from \(B_1\) to \(B_2\). The sequence \((\Pi_{AR}, \Pi_{BR})\) is a reversible standard weight process for \(CR\) from \(C_1\) to \(C_2\), in which the energy change of \(R\) is the sum of the energy changes in the constituent processes \(\Pi_{AR}\) and \(\Pi_{BR}\), i.e., \((\Delta E_{\text{rev}}^{A})_{C_1C_2} = (\Delta E_{\text{rev}}^{A})_{A_1A_2} + (\Delta E_{\text{rev}}^{A})_{B_1B_2}\). Therefore,

\[
\frac{(\Delta E_{\text{rev}}^{A})_{C_1C_2}}{T_R} = \frac{(\Delta E_{\text{rev}}^{A})_{A_1A_2}}{T_R} + \frac{(\Delta E_{\text{rev}}^{A})_{B_1B_2}}{T_R} \tag{18}
\]

Equation (18) and the definition of entropy (15) yield Eq. (17).

Comment. As a consequence of Theorem 4, if the values of entropy are chosen so that they are additive in the reference states, entropy results as an additive property.

Theorem 5. Let \((A_1, A_2)\) be any pair of states of a system \(A\) and let \(R\) be a thermal reservoir with temperature \(T_R\). Let \(\Pi_{AR\text{rev}}\) be any irreversible standard weight process for \(AR\) from \(A_1\) to \(A_2\) and let \((\Delta E_{\text{irr}}^{A})_{A_1A_2}\) be the energy change of \(R\) in this process. Then

\[
-\frac{(\Delta E_{\text{irr}}^{A})_{A_1A_2}}{T_R} < S_2^A - S_1^A. \tag{19}
\]

Proof. Let \(\Pi_{A_1R}\) be any reversible standard weight process for \(AR\) from \(A_1\) to \(A_2\) and let \((\Delta E_{\text{rev}}^{A})_{A_1A_2}\) be the energy change of \(R\) in this process. On account of Theorem 2,

\[
(\Delta E_{\text{rev}}^{A})_{A_1A_2} < (\Delta E_{\text{irr}}^{A})_{A_1A_2}. \tag{20}
\]

Since \(T_R\) is positive, from Eqs. (20) and (15) one obtains

\[
-\frac{(\Delta E_{\text{irr}}^{A})_{A_1A_2}}{T_R} < -\frac{(\Delta E_{\text{rev}}^{A})_{A_1A_2}}{T_R} = S_2^A - S_1^A. \tag{21}
\]

Theorem 6. Principle of entropy nondecrease. Let \((A_1, A_2)\) be a pair of states of a system \(A\) and let \((A_1 \rightarrow A_2)_W\) be any weight process for \(A\) from \(A_1\) to \(A_2\). Then, the entropy difference \(S_2^A - S_1^A\) is equal to zero if and only if the weight process is reversible; it is strictly positive if and only if the weight process is irreversible.

Proof. If \((A_1 \rightarrow A_2)_W\) is reversible, then it is a special case of a reversible standard weight process for \(AR\) in which the initial stable equilibrium state of \(R\) does not change. Therefore, \((\Delta E_{\text{rev}}^{A})_{A_1A_2} = 0\) and by applying the definition of entropy, Eq. (15), one obtains

\[
S_2^A - S_1^A = -\frac{(\Delta E_{\text{rev}}^{A})_{A_1A_2}}{T_R} = 0. \tag{22}
\]

If \((A_1 \rightarrow A_2)_W\) is irreversible, then it is a special case of an irreversible standard weight process for \(AR\) in which the initial stable equilibrium state of \(R\) does not change. Therefore, \((\Delta E_{\text{rev}}^{A})_{A_1A_2} = 0\) and Equation (19) yields

\[
S_2^A - S_1^A > -\frac{(\Delta E_{\text{irr}}^{A})_{A_1A_2}}{T_R} = 0. \tag{23}
\]

Moreover, if a weight process \((A_1 \rightarrow A_2)_W\) for \(A\) is such that \(S_2^A - S_1^A > 0\), then the process must be reversible, because we just proved that for any irreversible weight process \(S_2^A - S_1^A > 0\), if a weight process \((A_1 \rightarrow A_2)_W\) for \(A\) is such that \(S_2^A - S_1^A > 0\), then the process must be irreversible, because we just proved that for any reversible weight process \(S_2^A - S_1^A = 0\).

5. Conclusions

A general definition of thermodynamic entropy has been presented, based on operative definitions of all the concepts employed in the treatment, designed to provide a clarifying and useful, complete and coherent, minimal but general, rigorous logical framework suitable for unambiguous fundamental discussions on Second Law implications.

Operative definitions of system, state, isolated system, separable system, environment of a system, process and system uncorrelated from its environment have been stated, which are valid also in the presence of internal semipermeable walls, reaction mechanisms and external force fields. The concepts of heat and of quasistatic process are never mentioned, so that the treatment holds also for nonequilibrium states, both for macroscopic and few particles systems.

A definition of thermal reservoir less restrictive than in previous treatments has been adopted: it is fulfilled by any single-constituent simple system contained in a fixed region of space, provided that the energy values are restricted to a
suitable finite range. The proof that entropy is a property of the system has been completed by a new explicit proof that the entropy difference between two states of a system is independent of the initial state of the thermal reservoir chosen to measure it.

The definition of a reversible process has been given with reference to a given scenario, i.e., the largest isolated system whose subsystems are available for interaction; thus, the operativity of the definition is improved and the treatment becomes also more explicitly compatible with old (see, e.g., Beretta et al., 1984, Hatsopoulos and Beretta, 2008) and recent (see, e.g., Goldstein et al., 2006, Bennett 2008, Lloyd 2008, Maccone 2009) interpretations of irreversibility in the quantum theoretical framework.

Acknowledgments

G.P. Beretta gratefully acknowledges the Cariplo–UniBS–MIT-MechE faculty exchange program co-sponsored by UniBS and the CARIPLO Foundation, Italy, under grant 2008-2290.

References


Fermi, E., 1937, Thermodynamics, Prentice-Hall.


