Lecture notes in Thermal Philosophy: Entropy, Demon and the Direction of Time (BA Course)

Gábor Hofer-Szabó[*]

"[Thermodynamics] ...is the only theory of universal content concerning which I am convinced that, within the framework of the applicability of its basic concepts, it will never be overthrown."
(Einstein)

"Every mathematician knows it is impossible to understand an elementary course in thermodynamics."
(Arnold)

Contents

---

[*] Research Center for the Humanities, Budapest, email: szabo.gabor@btk.mta.hu

1 A brief history of thermodynamics

1. Pressure

Aristotle took hot and cold as the two of the four fundamental opposite qualities and held that nature contains no void (*horror vacui* thesis). This Aristotelian tradition has subsequently been demolished during the 17th century.

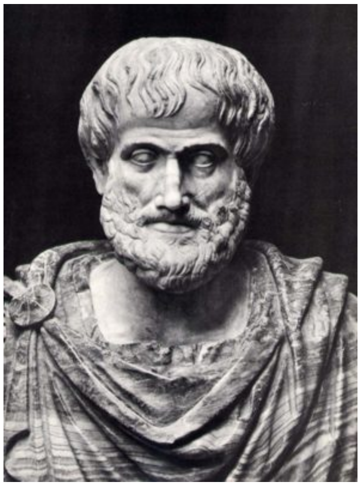


Figure 1: Aristotle, 4th century B.C.

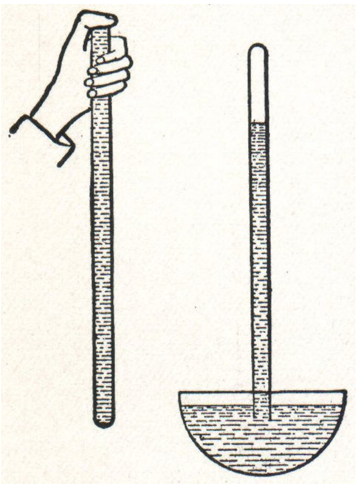


Figure 2: Torricelli's void experiment (1643)

The question of void became the center of interest when it was observed that pumps are able to raise water only up to roughly 10 m. The first to explain this fact by the air pressure was Beeckman in 1618. The *experimentum crucis* rebutting the horror vacui thesis was performed by Galileo's pupil Evangelista Torricelli in 1643.

The following decades brought with it passionate disputes concerning the nature of the Torricelli void. Descartes still endorsed the Peripatetic view on the *plenum* of the Cosmos, but Pascal's careful experiments in the 1640s have proven that there is vacuum in the Torricelli void. One of Pascal's famous experiment was his "void in the void" experiment.

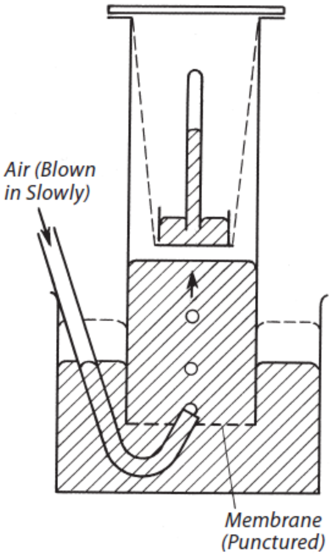


Figure 3: Pascal's void within a void experiment (1640s)

Figure 4: Perier's Puy-de-Dôme experiment (1648)

Pascal also devised an experiment, carried out by his brother-in-law Florin Perier on the mountain Puy-de-Dôme in 1648, to show that the height of the mercury column depended on the air pressure.

Otto von Guericke, the mayor of Magdeburg has succeeded in constructing an air pump which he used to perform many important experiments, among them his famous experiment with the hemispheres in 1654 in Resurgence in the presence of the emperor.

Figure 5: Guericke's experiment with the hemispheres (1654)

Robert Boyle, with his assistant Robert Hooke, has succeeded to build an air pump and studied the nature of the Torricelli void. He settled that light can travel through a vacuum, but sound cannot. By means of an U-shaped glass tube he refuted the idea that the pressure of air would not be able to raise a column of mercury to a height of 76 cm. He established the Boyle's law in 1662, according to which the product of the volume and pressure of a gas is constant, assuming constant temperature.

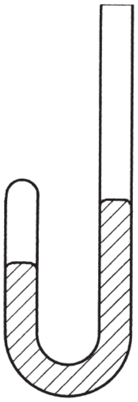


Figure 6: Boyle's U-shaped glass by means of which he established his law (1662)

2. Temperature. As for thermometry, Galileo was already using thermometers around 1600, but there were no standard fixed points. The epistemic question was how to judge whether a proposed fixed point is actually fixed in the absence of an already-trusted thermometer? Halley proposed for example deep caves as the lower fixed point; Joachim Dalencé proposed the melting point of butter as the upper fix point of the temperature scale; the Accademia del Cimento proposed the greatest summer heat; Newton proposed the human body temperature. By the middle of the eighteenth century, a consensus was emerging about using the boiling and freezing of water as the preferred fixed points, thanks to the Swedish Celsius.

Many early pioneers of thermometry, among them Anders Celsius, used „upside-down" thermometric scale. They may have thinking about measuring the degrees of cold than degrees of heat. The Aristotelian tradition regarded cold as real as heat. The mechanical philosophy of the seventeenth century understood heat as a motion and cold as the lack of it, but it did not plainly rule out cold as a substance. According to Francis Bacon, heat was a particular type of expansive motion and cold was a similar type of contractive motion. According to Pierre Gassendi, heat is caused by „caloric atoms" and cold is caused by „frigorific atoms" whose angular shapes made them damping down the motion of atoms. Marc-August Pictet, Genevan physicist and statesman, in a striking experiment seemed to confirm the reality of cold.

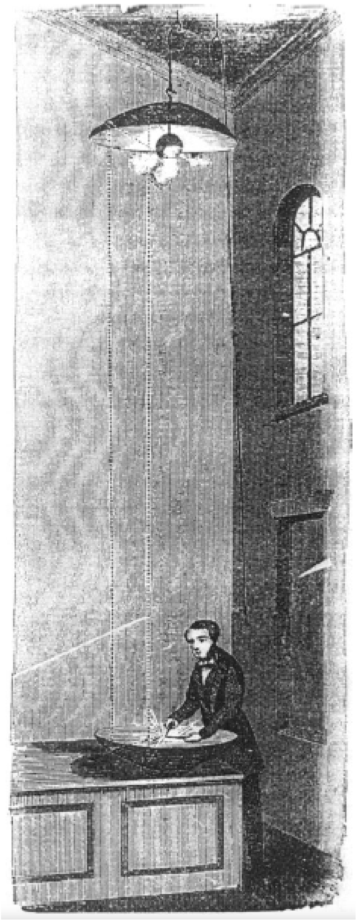


Figure 7: Pictet's experiment seemingly confirming the reality of cold (1791)

Scientists of the eighteenth century were realists in the sense that they believed in the existence of an objective property called temperature and persisted in wanting to know how to measure its true values. However, it turned out that different fluids expand in different ways. Three materials became especially significant: atmospheric air, mercury (quicksilver), and ethyl alcohol (spirit of wine). The method to decide which substance expands uniformly, one used the method of mixtures: mixing equal amount of freezing and boiling water the common temperature should be 50 C. De Luc, a proponent of the method, found mercury to be the best substance. However, the method of mixture assumed that the specific heat of substances does not depend on the temperature. This assumption was, however, challenged by the caloric theory.

The ruling theory of heat in the late 18th century was the *caloric theory* of heat advocated by Anton Lavoisier, Joseph Black, Joseph Fourier and William Irvine. It replaced the alchemical phlogiston theory of of combustion in the 17th century. The core of the caloric theory was the postulation of caloric, a material substance that was seen as a subtle, all-penetrating and weightless substance that was attracted to ordinary matter but self-repulsive (therefore elastic).

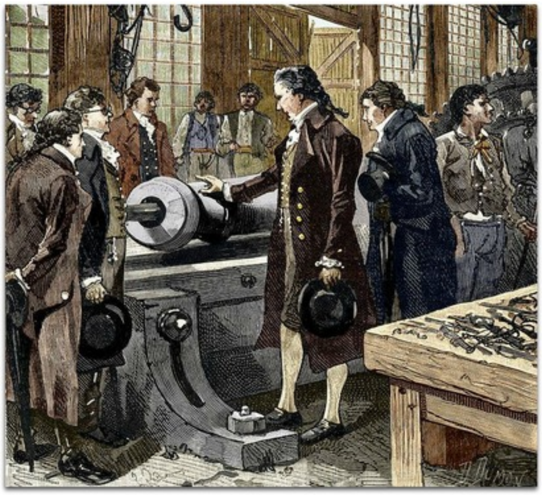


Figure 8: Rumford's cannon boring experiment (1798)

The caloric theory of heat has been disproved by Count Rumford's (Benjamin Thompson). Measuring the latent heat he was able to prove that the weight of the caloricum is practically zero. The weakest point of the caloric theory was its inability to provide a plausible explanation for the creation of heat through friction. His cannon boring experiment was a death-blow to the substance theory. In this experiment Rumford boiled water using the heat generated from the friction of continuous boring of a cannon. He made water boil in 2.5-hours time, without the use of fire.

Rumford's experiment inspired Joule to measure the equivalence of heat and work in a much more precise way. He devised an apparatus in which the work of a falling weight is converted into the heat of agitation in the water. Based on this experiment he formulated the principle of conservation of energy (also formulated by Julius Mayer earlier).

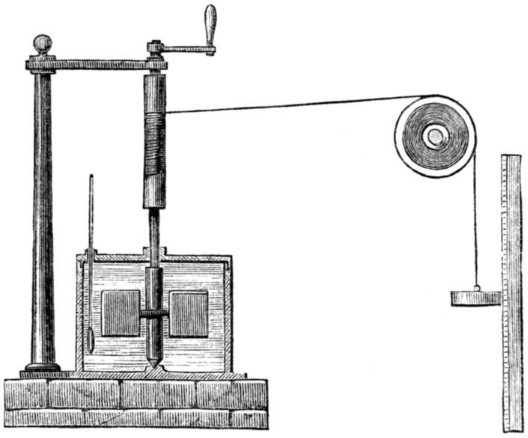


Figure 9: Joule's experiment (1843)

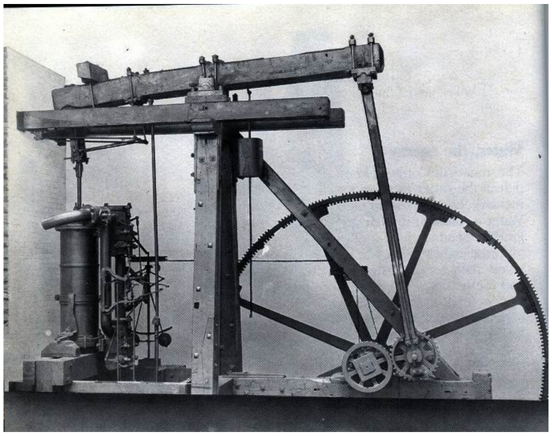


Figure 10: Watt's steam engine (1859)

Industrial Revolution has brought with it a widespread interest in the theory of heat engines which at that time were mostly steam engines and initially were used to pump water in order to drain deep mines. A decisive innovation of the steam engine was made by James Watt. He also invented the „expansive principle" that is the supply of steam can be cut off earlier because of the adiabatic expansion of the steam.

The calorist Sadi Carnot set himself the task to understand the efficiency of the heat engine using the analogy of a water wheel ($Q \sim$ mass, $T \sim$ height). He defined it as the ratio of the work performed and the amount of caloric passed through the engine:

$$\eta = \frac{Q(T_h - T_c)}{Q\,T_h} = 1 - \frac{T_c}{T_h} \leqslant 1$$

Since for the Carnot cycle

$$\frac{Q_c}{Q_h} = \frac{T_c}{T_h}$$

his definition based on an incorrect analogy has proved to be correct. His abstract heat engine was operating with water-steam mixture enclosed in a cylinder (rather than evacuated, as in Watt engine). The graphic representation of the Carnot cycle originates from Clapeyron in 1834.

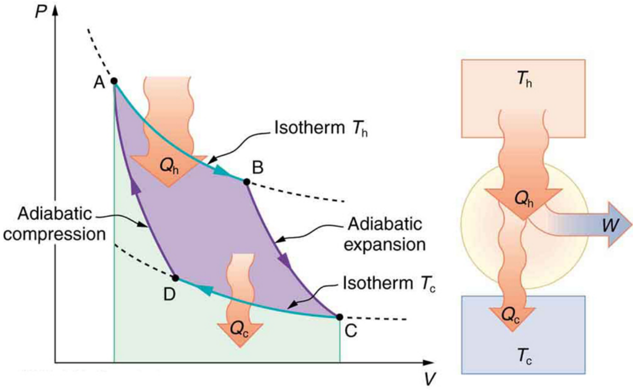


Figure 11: Carnot cycle (1824)

The kinetic theory of thermodynamics had many precursors. Daniel Bernoulli was the first to explain pressure as arising from the impact of molecular collision and showed that it was proportional to the *vis viva* (mv^2), but he did not identify *vis viva* with temperature. The idea that temperature is proportional to the molecular velocity (but not the square of velocity) came from John Herapath. John James Waterston, Scottish engineer was the first to identify temperature with mv^2 in 1843.

3. A brief chronology:

4th c., B.C. Aristotle's legacy: (i) Cold and hot are fundamental opposite qualities. (ii) There is no void.

1603 Galileo: air thermometer

1643 Evangelista Torricelli: vacuum experiment

1640s Blaise Pascal: void within a void experiment

1662 Robert Boyle's law (Boyle-Mariotte law)

1663 Otto von Guericke: Magdeburg hemispheres

1695 Guillaume Armontons: first prediction of the absolute zero temperature by extrapolating the pressure-temperature relation

1710s Daniel Gabriel Fahrenheit: mercury thermometer

c. 1740 Anders Celsius: centigrade thermometer

Late 18th c. Anton Lavoisier, Joseph Black: caloric theory of heat

1791 Marc-August Pictet: reality of cold?

1798 Count Rumford (Benjamin Thompson): cannon boring experiment

1799 Sir Humphry Davy: ice-rubbing experiments

1760-1840 Industrial revolution

1802 Joseph Louis Gay-Lussac's law

1824 Sadi Carnot: Carnot cycle

1842 Julius Robert Mayer: principle of conservation of energy.

1843 James Prescott Joule: equivalence of work and heat;

John James Waterston: identification of temperature with *vis viva* (mv^2)

1848 Lord Kelvin (William Thomson): absolute temperature scale based on the Carnot cycle

1850 Rudolf J. Clausius: first and second laws of thermodynamics; introduction of the concept of internal energy

1859 James Watt: steam engine

1875 Josiah Willard Gibbs: general formulation of thermodynamics

1897 Max Planck: new formulation of the second law of thermodynamics

1909 Constantin Carathéodory: axiomatic thermodynamics

Notes and further readings

2 Thermodynamics in a nutshell

4. Basic concepts

- Methodological starting point: fundamental laws should be independent of any hypothesis about the microscopic constitution of matter (Reason: positivist-empirical philosophy, multitude of microphysical hypotheses)
- Thermodynamic state: a small number of parameters describing the system's macroscopic state
    - Intensive and extensive parameters
- Equations of state: $f(p, V, T, n) = 0$
    - Ideal gas law: $pV = nRT$
- Temperature. To introduce *empirical* temperature choose a system with a convenient property x (length, pressure) and take temperature θ to be a linear function of x: $\theta(x) = ax + b$. Different θs need not linearly related. For the mercury thermometer x is the length L of the mercury column. To adjust the coefficients a and b, take two fixed points: the temperature of the boiling water θ_b and the temperature of the melting ice θ_m. For a centigrade thermometer set $\theta_b - \theta_m = 100$. Setting also $\theta_m = 0$ one obtains

  $$\theta(L) = \frac{100(L - L_m)}{L_b - L_m}$$

  For gas thermometers x is the pressure p of the gas and V and n are kept constant. Taking again a centigrade thermometer, one obtains:

  $$\theta(p) = \frac{100\,p}{p_b - p_m} + b$$

  Now, for low pressure the difference in θ between the different gases becomes small, that is, for different V, n and for different sorts of gases the above straight lines all meet at $p = 0$. Taking $\theta(p = 0) = 0$, the constant b will be 0. Substituting p_m in place of p we obtain $\theta_m = 273.16$. (For a thermometer of double V and same n, the measured p will be the half, but since it appears both in the nominator and denominator it drops out.)

  The reason for this empirical fact is that real gases at low pressure are close to ideal gases for which $T \sim p$. However, T is not the empirical temperature but the *thermodynamic* temperature (see later). That is we are lucky: gas thermometry is a good method for determining thermodynamic temperature.
- Thermodynamic equilibrium: thermodynamic coordinates are constant over time

- Processes:
  - Quasi-static process: the system is in equilibrium throughout the process ('Moving a cup of tea filled to the brim without spilling')
  - Reversible process: if the initial state of the system and the environment can be recovered from the final state (frictionless piston). A reversible process gives the maximum amount of work that a system can produce
  - Adiabatic process: there is no heat exchange
- Relations:
  - Reversible $\Longrightarrow$ quasi-static; but quasi-static $\not\Longrightarrow$ reversible (piston subject to friction)
  - Reversible $\not\Longrightarrow$ adiabatic (frictionless piston); adiabatic $\not\Longrightarrow$ reversible (free expansion)
  - Reversible $\neq$ time reversal invariant

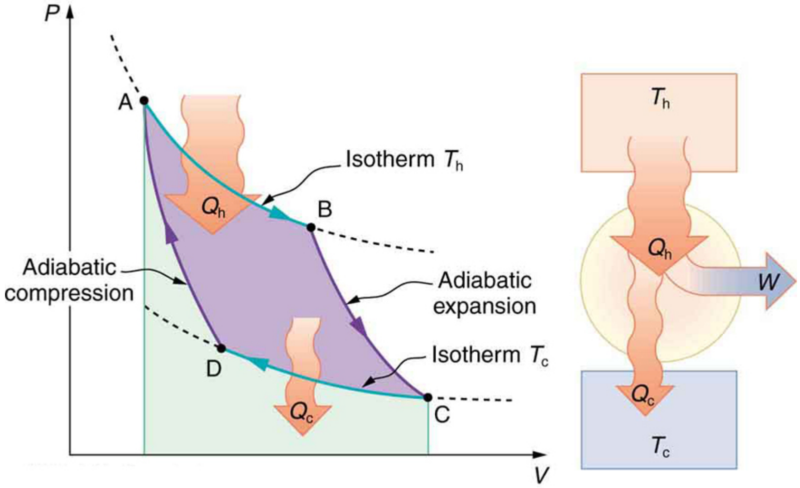
5. Laws

- 'Minus First' Law (Uffink): there is a spontaneous tendency to thermodynamic equilibrium
- Zeroth Law: 'is in thermal equilibrium with' is a transitive relation (necessary for defining temperature)
- First Law: $\Delta U = Q + W$
- ΔU is a function of state but Q and W are not. (The division of ΔU into Q and W is path dependent)
- Second Law:
  - Clausius' Principle: It is impossible to perform a cyclic process which has no other result than that heat is absorbed from a reservoir with a low temperature and emitted into a reservoir with a higher temperature.
  - Kelvin's Principle: It is impossible to perform a cyclic process with no other result than that heat is absorbed from a reservoir and work is performed.

  (If temperature is positive,) they are equivalent (Prove after the Carnot cycle!)

6. Carnot cycle and entropy

- Carnot cycle:



  - Efficiency of heat engine: $\eta = \frac{W}{Q_2} = 1 - \frac{Q_1}{Q_2} \leqslant 1$
  - Coefficiency of refrigerator: $\kappa = \frac{Q_1}{W} = \frac{Q_1}{Q_2 - Q_1} = \frac{1}{\eta} - 1 \geqslant 0$

- Carnot Principle:

  (i) All reversible heat engines operating between two heat reservoirs have the same efficiency.

  (ii) No heat engines between two given heat reservoirs can be more efficient than a Carnot engine.

- The Carnot Principle is the consequence of the Second Law

- Thermodynamic temperature. Since the efficiency of the reversible Carnot cycle depends only on the heat reservoirs and the only relevant property of the heat reservoirs is the temperature, one can define thermodynamic temperature *via* the efficiency of the Carnot cycle (first employed by Lord Kelvin). Define thermodynamic temperature as:

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

  For a Carnot cycle between between heat reservoirs of the boiling water and melting ice the efficiency is 0.268. Hence, taking

$$1 - \frac{T_m}{T_b} = 1 - \frac{Q_m}{Q_b} = \eta = 0.268 \quad \text{and} \quad T_b - T_m = 100$$

one gets $T_m = 273.16$. For an unknown temperature T just perform a Carnot cycle between T and, say, T_m, then measure the efficiency η and define T *via*

$$\frac{T_m}{T} = 1 - \eta$$

- Entropy
    - Efficiency for ideal gas: $1 - \frac{Q_1}{Q_2} = 1 - \frac{T_1}{T_2}$
    - Therefore, from the (reversible) Carnot principle, for any system: $-\frac{Q_1}{T_1} + \frac{Q_2}{T_2} = 0$
    - Generally: $\sum_i \frac{\pm Q_i}{T_i} \quad \longrightarrow \quad \oint \frac{dQ}{T}$
    - Introduction of a function of state: entropy: $S(A) := \int_0^A \frac{dQ}{T}$
    - For reversible process: $S(B) - S(A) = \int_A^B \frac{dQ}{T}$
    - For irreversible process (from the Second Law): $S(B) - S(A) > \int_A^B \frac{dQ}{T}$
- Second Law: $S(B) - S(A) \geqslant \int_A^B \frac{dQ}{T}$
    - The Second Law is exceptionless
    - Entropy is a property of an individual system
    - Entropy does not fluctuate
- For adiabatic ($\Delta Q = 0$) processes: $\Delta S \geqslant 0$
    - Reversible adiabatic expansion: $\Delta Q = 0$ therefore $\Delta S = 0$
    - Heat conduction (irreversible process): $\Delta S = \Delta S_1 + \Delta S_2 = \frac{\Delta Q}{T_1} - \frac{\Delta Q}{T_2} = \Delta Q(\frac{1}{T_1} - \frac{1}{T_2}) > 0$
    - Free expansion (irreversible process): $\Delta S = nR \log \frac{V_2}{V_1} > 0$

Notes and further readings

3 Classical mechanics

7. Hamiltonian mechanics. Classical mechanics (CM) can be treated in many different ways. The treatment best suited to the purposes of statistical mechanics (SM) is Hamiltonian mechanics.

In CM a system of n point particles is represented by a point $x = (q, p)$ in the *phase space* Γ of the system where the components q_i and p_i stand for the $3n$ coordinates and momenta of the particles, respectively. Occasionally, we also will use the notation q_a and p_a with the index $a = 1 \dots n$ running over the particles and not the degrees of freedom. The phase space is endowed with a normed Lebesgue (volume) measure μ. The time evolution of the system is given by the Hamilton equations:

$$\dot{q}_i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial q_i} \tag{1}$$

where $H = H(q_i, p_i, t)$ is the Hamiltonian of the system. The *trajectories* of the particles are solutions

$$x : \mathbb{R} \to \Gamma; \quad t \mapsto x(t)$$

of the Hamilton equations (1). The Hamilton equations define a *Hamiltonian flow*, a one parameter group of transformations mapping the phase space onto itself for any t. An important theorem of Hamiltonian mechanics is

Liouville's theorem. The Lebesgue measure is invariant under the Hamiltonian flow: for any t and $A \in \mathcal{F}$, where $\mathcal{F}$ is the Lebesgue measurable subsets of Γ: $\mu(\phi_t(A)) = \mu(A)$.

The triple (Γ, μ, ϕ_t) is called a *Hamiltonian dynamical system.*

The conservation of energy confines the system to the energy hypersurface

$$\Gamma_E := \{x \in \Gamma \mid H(x) = E\}$$

and induces[1] a measure μ_E on Γ_E which also is invariant: $\mu_E(\phi_t(A_E)) = \mu_E(A_E)$ for any t and $A_E \in \mathcal{F}_E$.

---

[1] By

$$\mu_E(A_E) := \int_{A_E} \frac{d\sigma_E}{||\nabla H||}$$

where $d\sigma_E$ is the surface element on Γ_E and

$$||\nabla H|| := \left[\sum_{i=1}^{n} \left(\frac{\partial H}{\partial p_i}\right)^2 + \left(\frac{\partial H}{\partial q_i}\right)^2\right]^{\frac{1}{2}}$$

8. Harmonic oscillator. A simple example of Hamilton mechanics is the harmonic oscillator (spring). The Hamiltonian of the spring is

$$H = \frac{1}{2m}p^2 + \frac{1}{2}kq^2$$

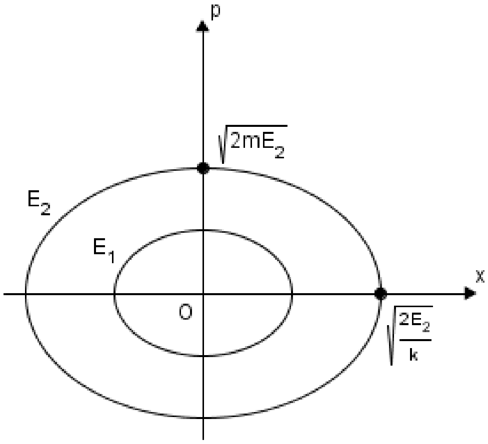
and the equations of motions are

$$\dot{q} = \frac{p}{m}, \quad \dot{p} = -kq \tag{2}$$

Solving (2) one obtains the trajectories which are the following ellipses:

$$x(t) = (q(t) = A_1 \sin(\omega t + \phi_0), \, p(t) = A_2 \cos(\omega t + \phi_0))$$

with $A_1 = \sqrt{\frac{2E}{k}}, A_2 = \sqrt{2mE}$ and $\omega = \sqrt{\frac{k}{m}}$.



One can show that the Hamilton flow leaves the volume measure $d\mu = dq\,dp$ invariant and also the line measure μ_E.

Notes and further readings

Comprehensive presentations of CM include Arnold (1978), Abraham and Marsden (1980) and Butterfield (2007).

4 Time reversal invariance and irreversibility

9. Symmetries. A physical system has a *symmetry* if there is a group action of the group G on $\mathbb{R} \times \Gamma$

$$\alpha : G \times \mathbb{R} \times \Gamma \to \mathbb{R} \times \Gamma; \quad (g, x(t)) \mapsto \alpha_g(x(t))$$

such that for any solution $x(t)$ of the equations of motion of the system, $\alpha_g(x(t))$ is also a solution. A sufficient condition for this is that the equations of motion remain invariant under the group action or that the Hamiltonian of the system remains invariant (up to a constant).

Spatial translation: The symmetry group G is the N-fold direct sum of the translation group $T(3)$, and for any $a = 1 \dots n$

$$\alpha_{\Delta q_a}\big((q_a(t), p_a(t))\big) = (q_a(t) + \Delta q_a, p_a(t))$$

where Δq_a is a constant spatial vector. Spatial translation is *not* a symmetry of the harmonic oscillator, since

$$x(t) = \big(q(t) + \Delta q,\, p(t)\big)$$

is not a solution of (2). (Substitute!) Equivalently,

$$\dot{q} = \frac{p}{m},\ \dot{p} = -k(q + \Delta q)$$

is not of the same form as (2). The harmonic oscillator is not spatial translation invariant. Generally, if H is spatial translation invariant, then the system is spatial translation invariant.

Spatial rotation: The symmetry group G is the N-fold direct sum of the rotation group $SO(3)$, and for any $a = 1 \dots n$

$$\alpha_R\big((q_a(t), p_a(t))\big) = (Rq_a(t), Rp_a(t))$$

where $R \in SO(3)$.

Spatial inversion: The symmetry group is $\mathbb{Z}_2$, and

$$\alpha_-\big((q(t), p(t))\big) = (-q(t), -p(t))$$

Spatial inversion *is* a symmetry of the harmonic oscillator, since

$$x(t) = \big(-q(t),\, -p(t)\big)$$

is a solution. The harmonic oscillator is spatial inversion invariant. Generally, if H is invariant under spatial inversion, then also the Hamilton equations, since:

$$\frac{\partial(-q_i)}{\partial t} = \frac{\partial H}{\partial(-p_i)}, \quad \frac{\partial(-p_i)}{\partial t} = -\frac{\partial H}{\partial(-q_i)}$$

is of the same form as (2).

Time translation: The symmetry group is again $T(1)$, and

$$\alpha_{\Delta t}\big((x(t))\big) = (x(t+\Delta t))$$

where Δt is a constant scalar. Time translation is also a symmetry of the harmonic oscillator, since

$$x(t) = \big(q(t+\Delta t),\, p(t+\Delta t)\big)$$

is a solution of (2). The harmonic oscillator is time translation invariant. Generally, if H is autonomous, $H \neq H(t)$, then the system is time translation invariant.

Noether's theorems connect symmetries and conservation laws in the following way:

- If H is autonomous, $H \neq H(t)$, then

  $$\frac{dH}{dt} = \frac{\partial H}{\partial t} + \sum_i \left(\frac{\partial H}{\partial q_i}\dot{q}_i + \frac{\partial H}{\partial p_i}\dot{p}_i \right) = \frac{\partial H}{\partial t} + \sum_i \left(\frac{\partial H}{\partial q_i}\frac{\partial H}{\partial p_i} - \frac{\partial H}{\partial p_i}\frac{\partial H}{\partial q_i} \right) = 0$$

  hence the energy $E = H$ is a constant of motion.

- If $H \neq H(q_i)$, then

  $$\dot{p}_i = -\frac{\partial H}{\partial q_i} = 0$$

  hence the momentum p_i is a constant of motion.

Finally, let us turn to time reversal.

10. Time reversal. The symmetry group is again $\mathbb{Z}_2$, and

$$\alpha_-\big((q(t), p(t))\big) = (q(-t), -p(-t))$$

Time reversal *is* a symmetry of the harmonic oscillator. Generally, if $H \neq H(t)$ and $H(q_i, p_i) = H(q_i, -p_i)$ then the system is time reversal invariant, since

$$\frac{\partial q_i}{\partial(-t)} = \frac{\partial H(q_i, -p_i)}{\partial(-p_i)}, \quad \frac{\partial(-p_i)}{\partial(-t)} = -\frac{\partial H(q_i, -p_i)}{\partial q_i}$$

are equivalent to (1). All systems with Hamiltonian of the form

$$H = \sum_a \frac{1}{2m_a} p_a^2 + U(q_a)$$

are of this kind. Finally, note that conservation of energy

$$\frac{dH}{dt} = \frac{\partial H}{\partial t} = 0$$

is *not* enough for time reversal invariance.

This leads us to the definition of time reversal invariance. Let $x(t)$ be a trajectory in the phase space of a Hamiltonian system and let $T(x(t))$ denote the time reversal of $x(t)$. Let S denote the set of solutions that is all the trajectories which are solutions for the equations of motion of the theory.

Time reversal invariance. A Hamiltonian system is called *time reversal invariant* if from $x(t) \in S$ it follows that $T(x(t)) \in S$. In other words, the time reversed of any solution (process, possible world) is a solution again.

A central question in physics is to check whether a given physical theory, not necessarily CM is time reversal invariant or not. In order to decide the question, first one needs to define the time reversal transformation in every single theory. If we concentrate only on its action on the phase space, then a time reversal transformation T is an involution ($T^2 = id$). In classical mechanics it is defined as: $T(q, p) = (q, -p)$, in classical electrodynamics as: $T(E, B, \rho, j) = (E, -B, \rho, -j)$, in quantum mechanics as: $T(\psi) = \psi^*$ and in thermodynamics as: $T = id$.

11. Albert vs. Earman and Malament on the physical states. Albert characterizes the time reversal simply as $T(s(t)) = s(-t)$ and takes the state $s(t)$ to be different from how it is characterized in CM, namely $x(t)$. He takes $s(t)$ to be an *instantaneous state* of the system defined as follows:

(i) $s(t)$ must be *genuinely instantaneous*, that is the state descriptions at different times must be logically, conceptually, and metaphysically independent.

(ii) It must be *complete*, that is all of the physical facts about the world can be read off the full set of state descriptions.

Albert identifies $s(t)$ with $q(t)$ and not with $(q(t), p(t))$, and argues that the notion of velocity is not logically independent form the state of system at other times (fixing the position at other times fixes the velocity now). He takes CM time reversal invariance in both senses:

$$T(q(t)) = q(-t) \tag{3}$$
$$T(q(t), p(t)) = (q(-t), -p(-t)) \tag{4}$$

But he takes (4) to be a consequence of (3): if the path is reversed the velocities change sign. However, in classical electrodynamics he takes $T(E, B, \rho, j) = (E, -B, \rho, -j)$ to be unjustified, since "magnetic fields . . . are not the rates of change of anything" (Albert, 2002, 20). Therefore, classical electrodynamics is not time reversal invariant. Moreover, "none of the fundamental physical theories that anybody has taken seriously throughout the past century and a half is . . . invariant under time-reversal." (Albert, 2002, 15)

Earman defends the traditional view both in the physical state description and also in the time reversal invariance of the different physical theories. He says that physics takes instead of (ii) another completeness criterion:

(ii') All of the physical facts about the world *at the time* in question can be read off the state description *at that time*.

The simply motivation for this completeness criterion is that one would like to distinguish the stationary and the moving arrow of Zeno. Physicists also want *dynamical completeness* of state descriptions since do not want to preclude *ab initio* the possibility of Laplacian determinism.

Two remarks on time reversal invariance:

1. *Malament's theorem.* Let V^μ be a unit timelike vector field and let the four-vector electric field E^μ relative to V^μ be defined from the Maxwell tensor $F^{\mu\nu}$ by $F^\mu_\nu V^\nu$. If $(TE)^\mu = E^\mu$ and if $(TE)^\mu$ is defined as $(TF)^\mu_\nu (TV)^\nu$, then the magnetic field must obey the standard transformation law.

2. Time reversal invariance does not mean that we cannot *distinguish between future and past* in the sense that for any solution $x(t)$: $T(x(t)) = x(t)$. Time reversal invariance states only that the *set* of solutions is invariant, not the solutions individually. This latter holds only if the laws are deterministic and $(Ts)_t = s_t$ for a certain time t, which holds only if the velocities are zero.

   "This is just an instance of the more general fact that contingent conditions don't have to exhibit the symmetries of the laws that govern them, e.g. spatial translation invariance of the laws doesn't mean that the matter content of space has to be arranged homogeneously." (Earman, 2002, 254)

12. Reversibility. The differences between time reversal invariance and reversibility are the following:

1. While time reversal invariance is attributed to a theory, reversibility is a feature of a process: it means the complete recoverability of the initial state of the system together with its environment.

2. If a reversible process can be undone by another process, then this new process need not be the original process retraced step by step in the reverse order.

3. 'Complete recovery' of the initial state involves, not only the system itself, but also its environment.

Therefore, time-reversal non-invariance and irreversibility are logically independent notions.

Notes and further readings

5 Statistical mechanics: the Boltzmann program

13. The aim of statistical mechanics (SM) is to give an account for the thermodynamic behavior of macroscopic systems in terms of the dynamical laws governing their microscopic constituents and some probabilistic assumptions. There are two subprojects:

Equilibrium theory: What the dynamical laws have to be to ensure that the macroscopic parameters remain constant? How can the values of macroscopic parameters like pressure and temperature be calculated on the basis of the microphysical description?

Non-equilibrium theory: How does a system approach equilibrium when left to itself in a non-equilibrium state and why? What is microphysical explanation of the fact that processes in the world are irreversible?

There are two approaches in SM: the Boltzmannian and the Gibbsian approach. We start with the Boltzmannian.

14. Basic notions. In classical SM the micromechanism underlying TD is represented by CM. At the thermodynamic level the system is characterized by *macrostates* M_i ($i = 1 \ldots m$). The basic assumption of the Boltzmannian approach is the macrostates supervene on microstates, that is to every microstate x there correspond exactly one macrostate, $M_i(x)$. Let Γ_{M_i} denote the *macroregion* associated to M_i on Γ_E:

$$\Gamma_{M_i} := \{x \in \Gamma_E \,|\, M(x) = M_i\} \tag{5}$$

$\{\Gamma_{M_i}\}_i$ forms a partition of Γ_E.

The Boltzmann entropy of the macrostate M_i is defined as:

$$S_B(M_i) := k \, \log \left(\mu_E(\Gamma_{M_i})\right) \tag{6}$$

where $k = 1,38 \cdot 10^{-23} \frac{J}{K}$ is the Boltzmann constant. The logarithmic function is usually justified by rendering entropy to be an *extensive* quantity which it is in TD. The *equilibrium macrostate*, M_{eq}, is, by definition, the macrostate for which the Boltzmann entropy is maximal. The aim of the Boltzmann program is to provide a probabilistic justification of the Second Law in the following form:

Statistical Second Law: If the Boltzmann entropy $S_B(t_1)$ of the system at time t_1 is below the maximal value, then it is highly probable that at a later time t_2 we will have: $S_B(t_2) \geqslant S_B(t_1)$.

Callender (2001) argues that TD should not be taken too seriously and irreversibility should be understood only in the sense of the Statistical Second Law.

Probabilities enter into SM in two different ways. The one is the *Proportionality Postulate*: the probability of a macrostate M_i is proportional to the measure of the corresponding macroregion:

$$p(M_i) := c \, \mu_E(\Gamma_{M_i}) \tag{7}$$

where c is a normalization constant. The other is the *Statistical Postulate*: the probability that the system's microstate lies in a region $A_E \in \mathcal{F}_E$, given its macrostate is M, is

$$\frac{\mu_E(A_E)}{\mu_E(\Gamma_M)}$$

15. Boltzmann's (1877) combinatorial argument. Represent the system now not by a point in the $6n$ dimensional Γ-space, but by n points in the 6 dimensional one-particle phase space, the μ-space. Coarse-grain the μ-space into k *cells* $(\omega_1, \omega_2 \ldots \omega_k)$ such that each cell is a 6 dimensional cubelet with equal volume $\delta\omega = \delta^3 x\, \delta^3 v$. The cells should be small enough compared to the system, but big enough to contain enough particles. There are three key notions in the combinatorial argument:

(i) *Microstate*: $x \in \Gamma$ or, equivalently, n points in the μ-space.

(ii) *Arrangement* (coarse-grained microstate): the specification of which of the n particles is in which cell.

(iii) *State distribution* $Z_{\{n_i\}} = (n_1, n_2 \ldots n_k)$: the specification of how many particles are in which cell. The number n_i is called the *occupation number* of cell ω_i and $\sum_{i=1}^{k} n_i = n$.

It is assumed that the macrostates of the system depend only on the number of particles in the different cells, hence state distributions are identified with *macrostates*. Macrostates supervene on arrangements and arrangements supervene on microstates: permutation of particles changes the arrangement but not the distribution; moving particles within the cells changes the microstate but not the arrangement. The number of arrangements in a state distribution $Z_{\{n_i\}}$ is

$$\frac{n!}{n_1!\, n_2!\, \ldots\, n_k!}$$

Each distribution (macrostate) corresponds to a well defined region in the Γ-space, the *macroregion*

$$\Gamma_{Z_{\{n_i\}}} := \{x \in \Gamma \,|\, Z = Z_{\{n_i\}}\} \tag{8}$$

Macroregions form a partition of Γ: $\{\Gamma_{Z_{\{n_i\}}}\}$. Since each cell corresponds to a region of equal volume $(\delta\omega)^n$ in Γ, therefore the volume of the macroregion $\Gamma_{Z_{\{n_i\}}}$ will be proportional to the number of arrangement in a given distribution:

$$\mu\left(\Gamma_{Z_{\{n_i\}}}\right) = \frac{n!}{n_1!\, n_2!\, \ldots\, n_k!}(\delta\omega)^n$$

The volume of the different macroregions vary enormously in size: if there are many particles in the same cell, then $\mu\left(\Gamma_{Z_{\{n_i\}}}\right)$ will be small; for evenly distributed arrangements

$\mu\left(\Gamma_{Z_{\{n_i\}}}\right)$ will be big. The *equilibrium macrostate* is defined to be the distribution corresponding to the largest macroregion (largest 'number' of different microstate). Which is that?

Boltzmann assumed that the energy E_i of a particle depends only on the cell ω_i the particle is in, that is he neglected interaction between the particles. Hence, the total energy of the system can be given as $\sum_{i=1}^{k} n_i E_i = E$. With the constraints on Z:

$$\sum_{i=1}^{k} n_i = n, \quad \sum_{i=1}^{k} n_i E_i = E$$

using the method of Lagrange multipliers and the Stirling formula, Boltzmann shows that the equilibrium macrostate is the one for which the occupation numbers are:

$$n_i = \alpha e^{-\beta E_i}$$

where α and β are constants. This is the discrete Maxwell-Boltzmann distribution.

16. Coarse-grained Boltzmann entropy. The combinatorial argument can also be expressed in terms of entropy. Since macroregions Γ_{M_i} in (5) are in Γ_E whereas the macroregions $\Gamma_{Z_{\{n_i\}}}$ in (8) of the combinatorial argument are in Γ, first we need to identify them by intersecting by the energy hypersurface:

$$\Gamma_{M_i} = \Gamma_{Z_{\{n_i\}}} \cap \Gamma_E \tag{9}$$

We also need to assume the $6n$ dimensional measure of $\Gamma_{Z_{\{n_i\}}}$ is proportional to the $6n-1$ dimensional measure of Γ_{M_i} for any i:

$$\mu_E\left(\Gamma_{M_i}\right) = c\,\mu\left(\Gamma_{Z_{\{n_i\}}}\right) \tag{10}$$

Hence the Boltzmann entropy (6) can be rewritten as:

$$S_B(M_i) = k\,\log\left(\mu_E(\Gamma_{M_i})\right) = k\,\log\left(c\,\mu\left(\Gamma_{Z_{\{n_i\}}}\right)\right) = k\,\log\left(\mu\left(\Gamma_{Z_{\{n_i\}}}\right)\right) + k\,\log c$$

Now, defining the coarse-grained Boltzmann entropy as

$$S_{B,c}\left(Z_{\{n_i\}}\right) := k \log\left(\mu\left(\Gamma_{Z_{\{n_i\}}}\right)\right) \tag{11}$$

Boltzmann entropy and coarse-grained Boltzmann entropy will differ only in a constant which does not matter. By means of the Stirling formula the coarse-grained Boltzmann entropy can be written as:

$$S_{B,c}\left(Z_{\{n_i\}}\right) = \sum_i n_i \log n_i + c'(n, \delta\omega) \tag{12}$$

Introducing the probability $p_i = \frac{n_i}{n}$ of finding a randomly chosen particle in cell ω_i we obtain:

$$S_{B,c}\left(Z_{\{n_i\}}\right) = \sum_i p_i \log p_i + c''(n, \delta\omega) \tag{13}$$

p_i is not to be confused by $p(M_i)$ defined in (7). The coarse-grained Boltzmann entropy is: (i) a measure for the number of arrangements compatible with a given macrostate, (ii) a measure of how much we can infer about the arrangement of the system on the basis of its macrostate. Again, equilibrium macrostate is, by definition, the macrostate with the largest entropy.

17. Remarks.

(i) In the combinatorial argument there is no mention of dynamics, collisions etc.

(ii) It works only for systems with no interaction. The energy of the particles depended only on its coarse-grained microstate and not on the state of other particles.

(iii) Is coarse-graining subjective or objective?

(iv) Taking energy cells the combinatorial argument fails to reproduce the Maxwell-Boltzmann distribution.

(v) Cells cannot be eliminated: in the $\delta\omega \to 0$ limit, the coarse-grained Boltzmann entropy diverges.

(vi) The definition of equilibrium has been changed: in TD equilibrium meant stationarity.

(vii) In what sense does the combinatorial argument explain the Statistical Second Law?

Notes and further readings

For the correct reductive relation between TD and SM see Callender (2001). In many respect we follow the exposition of SM of Uffink (2007) and Frigg (2008).

6 Ergodicity

18. Aims. In order to explain the Statistical Second Law it is not enough to say that the equilibrium macroregion is enormously (typically 10^n times) larger than the other regions. One also needs to know that the non-equilibrium regions are not dynamically detached from the equilibrium macroregion, that is the microstate of the system randomly wanders on the energy hypersurface. This is meant to be addressed by the ergodic programme.

19. Ergodicity. Suppose that the underlying dynamics of our thermodynamic system under investigation can be represented by a dynamical system $(\Gamma_E, \mu_E, \phi_t)$. Let $f : \Gamma_E \to \mathbb{R}$ be a *phase function* that is a Lebesgue integrable function on the accessible region of the phase space. Since measurement takes time, when measuring a quantity f what we actually obtain is the (infinite) *time average* of f:

$$\overline{f}(x) := \lim_{T\to\infty} \frac{1}{T} \int_0^T f(\phi_t x) dt \tag{14}$$

Taking the measurement time to be infinite is a strong idealization. Time average is empirically meaningful but hard to calculate. On the other hand we have another average of f, the *ensemble average:*

$$\langle f \rangle := \int_{\Gamma_E} f d\mu_E \tag{15}$$

which is easy to calculate but it is not clear how to interpret it. The aim of the ergodic programme is to relate the two averages. The following theorem claims that time averages exist:[2]

Birkhoff theorem. Let (Γ, μ, ϕ_t) be a dynamical system and let f be an integrable phase function. Then the time average $\overline{f}(x)$ (i) exists almost everywhere, (ii) is invariant: $\overline{f}(x_0) = \overline{f}(\phi_t x_0)$ for all t, and integrable: $\int_\Gamma \overline{f}(x_0) d\mu = \int_\Gamma f d\mu$.

Then ergodicity is defined as follows:

Ergodicity. A dynamical system (Γ, μ, ϕ_t) is *ergodic* iff for any measurable set $A \in \mathcal{F}$ such that $\mu(A) \neq 0$ and for all $x \in \Gamma$, except for a set of measure 0, it holds that $\{\phi_t(x)\} \cap A \neq \emptyset$ for some t.

The central theorem of the ergodic theory is the

Ergodic theorem. Let (Γ, μ, ϕ_t) be a dynamical system and let f be an integrable phase function. Then $\overline{f} = \langle f \rangle$ iff the system is ergodic.

The Ergodic theorem has three corollaries:

---

[2] From now on for the sake of simplicity I will use Γ instead of Γ_E and μ instead of μ_E but everything is understood to happen on the energy hypersurface.

(i) Let f be χ_A. Then $\mu(A) = \langle \chi_A \rangle = \overline{\chi}_A = \lim_{T\to\infty} \frac{1}{T} \int_0^T \chi_A(\phi_t x)dt$: the measure of a region is the fraction of time the system spends in the region: a frequentist justification of the volume measure μ.

(ii) Since $\overline{f} = \langle f \rangle$ and $\langle f \rangle$ does not depend on x, therefore neither $\overline{f}$ does. Therefore trajectories started from almost any point will come arbitrary close to any points (= will intersect any region of non-zero measure) on the energy hypersurface infinitely many times.

(iii) Let (Γ, μ, ϕ_t) be ergodic. Let μ' be a ϕ_t-invariant measure that is absolutely continuous with respect to μ (that is for any measurable $A \subseteq \Gamma$ if $\mu(A) = 0$, then $\mu'(A) = 0$). Then $\mu' = \mu$. (Loosely: if the system is ergodic, there is only one time invariant measure).

To sum up, if the system is ergodic then $\mu_E(\Gamma_{M_i})$ can be interpreted as a fraction of time (sojourn time) the system spends in macroregion Γ_{M_i}. Since the equilibrium macroregion is, by definition, the one with the largest μ_E-measure, it will also be the region in which the system spends most of its time. Hence, an ergodic system started from almost microstate will evolve into $\Gamma_{M_{eq}}$ and stays there except for some tiny fluctuations.

20. Ergodic hierarchy. There is an equivalent definition of ergodicity:

Decomposability. A dynamical system (Γ, μ, ϕ_t) is *decomposable* iff Γ can be partitioned into two (or more) invariant regions of non-zero measure, that is if there are $A, B \subseteq \Gamma$ such that $A \cap B = \emptyset$, $A \cup B = \Gamma$, $\mu(A) \neq 0 \neq \mu(B)$ and $\phi_t(A) \subseteq A$ and $\phi_t(B) \subseteq B$ for all t. A system is *indecomposable* (*metrically transitive*) if not decomposable.

In a decomposable system $\langle f \rangle \neq \overline{f}_A$: Let f be χ_A where A is an invariant region with $\mu(A) \neq 0, 1$. Then $\mu(A) = \langle \chi_A \rangle \neq \overline{\chi}_A = 1$.

There is another ergodic property which is a stronger property than ergodicity:

Mixing. A dynamical system (Γ, μ, ϕ_t) is *mixing* iff for all $A, B \in \mathcal{F}$: $\lim_{t\to\infty} \mu(\phi_t A \cap B) = \mu(A)\mu(B)$. Intuitively, mixing means that the dynamics homogeneously distributes the points of A over all measurable set B in the long run (cocktail analogy).

Mixing entails indecomposability: Let $B = A$ invariant subsets of Γ. Then $\lim_{t\to\infty} \mu(\phi_t A \cap A) = \mu(A) = \mu(A)\mu(A) \implies \mu(A) \in \{0, 1\}$. Indecomposability, however, does not entail mixing: 1-dimensional harmonic oscillator is indecomposable but not mixing: The energy hypersurface is the trajectory itself (an ellipse) and hence indecomposable but not mixing.

In studying deterministic random systems a whole hierarchy of ergodic systems was born (Bernoulli system $\implies$ K system $\implies$ mixing $\implies$ weak mixing, ergodicity = indecomposability).

21. Discussion. It is extremely difficult to prove that a system is ergodic, let alone mixing, etc. Ergodicity is proven for a system of $n \leqslant 4$ elastic hard balls moving in a cubic box and a system of $n \geqslant 2$ elastic hard balls moving in a k-torus (Sinai's ergodic hypothesis (1963) solved by Simányi, Szász). Ironically, it is known for many systems that they are *not* ergodic: Kac ring model; a system of n uncoupled anharmonic oscillators of identical mass; molecules oscillating in a solid; a system of non-interacting point particles. Moreover, the KAM-theorems[3] indicate that ergodicity cannot be a generic property in Hamiltonian systems. Since we observe thermodynamic behavior also in non-ergodic systems, ergodicity cannot provide the whole story in explaining the phenomena. Moreover, thermodynamic behavior is thought to be arising from a system consisting of a *large* number of particles. On the other hand, ergodicity identifies ensemble average with *infinite* time average which is an irrealistic assumption. (Moreover, why should observables be equated with time averages at all?) Finally, the Birkhoff theorem holds only *almost* everywhere. But what justifies the neglect of measure zero sets?

Notes and further readings

---

[3] KAM theorems show that in typical Hamiltonian systems the trajectories confined to an invariant set will continue to have that property even is perturbed.

7 Objections to the Boltzmann program

22. Historically there have been two major objections to the Boltzmann program: Loschmidt's Reversibility objection (*Umkehreinwand*, 1876) and Zermelo's Recurrence objection (*Wiederkehreinwand*, 1896). Both objections were devised to show that the irreversible character of TD is incompatible with the underlying theory of CM. Loschmidt's Reversibility objection was directed against Boltzmann's earlier work on SM in the 1870s which attempted to provide a non-statistical underpinning to TD. Boltzmann's (1877) combinatorial argument was a response to Loschmidt. Zermelo's Recurrence objection shed light on the role of initial conditions in the statistical arguments and highlighted what later became known as the Past Hypothesis, an assumption about the initial condition of the Universe.

23. The Reversibility objection has two premises:

1. CM is time reversal invariant: if $x(t)$ is a solution, then $T(x(t)) = (q(-t), -p(-t))$ also is a solution.
2. Marcostates are "time reversal invariant": for a macrostate M define the time reversed macrostate $T(M)$ by means of the time-reversed macroregion

   $$\Gamma_{T(M)} := \{x \in \Gamma \,|\, T(x) \in \Gamma_M\} \tag{16}$$

   Then according to the second premises: $T(M) = M$. Consequently, $S_B(M) = S_B(T(M))$.

As a consequence, for any trajectory (solution) $x(t)$ such that $x(t_1) \in M_1$, $x(t_2) \in M_2$, $t_1 < t_2$ and

$$S_B(M_1) < S_B(M_2)$$

there exist a time reversed trajectory $T(x(t))$ such that $-t_2 < -t_1$ and

$$S_B(T(M_2)) = S_B(M_2) > S_B(M_1) = S_B(T(M_1))$$

In other words, for any time segment for which the entropy is increasing there is another time segment for which the entropy is decreasing. This argument has a bite against the reduction of the original Second Law but not against the Statistical Second Law formulated above. Actually, Loschmidt's Reversibility objection was a motivation *for* the statistical reading of it. Boltzmann was not deeply moved by the argument; he responded laconically: "Go ahead and reverse the momenta!"

Now, though Loschmidt's Reversibility objection does not contradict to the Statistical Second Law, it still makes it difficult to understand *why* the Statistical Second Law holds true. Let $t_1 = 0$. Then for any trajectory starting from the macrostate M_1 at t_1 and ending up in a macrostate M_2 of higher entropy at a *later* time t_2, there is also a trajectory starting from the same macrostate M_2 at an *earlier* time $-t_2$ and ending up in a macrostate M_1 at time t_1. Hence, the measure $\mu_E(\Gamma_{M_1^+})$ of those microstates in M_1 for which the entropy is increasing in the next Δt time period is equal to the measure $\mu_E(\Gamma_{M_1^-})$ of those microstates in M_1 for which the entropy has been decreasing for the previous Δt period. Moreover, if "highly probable" in Statistical Second Law means that

$$\frac{\mu_E(\Gamma_{M_1^+})}{\mu_E(\Gamma_{M_1})} \approx 1$$

then it follows that most of the microstates in M_1 are "turning points" that is points such that the Boltzmann entropy on the adjacent trajectory increases in both forward and backward in time. And this is case with any non-equilibrium macrostate.

24. The Past Hypothesis is Boltzmann's response to the Reversibility objection. The Reversibility objection predicts an entropy increase for almost all macrostates in *both* time directions. This prediction is factually false since entropy of a typical thermodynamic system decreases as we go back in time. To resolve the contradiction between the Reversibility objection and empiricism, one needs to find an argument why statistical explanations cannot be used for retrodiction. One can then argue that physical laws, such as the Statistical Second Law, can account for physical phenomena only together with *initial conditions.* The proponents of this argument then posit a low entropy initial condition for the system under investigation. Hence, the entropy of the system cannot increase backward in time simply because the time period *before* the initial condition is not part of the process under investigation.

"Initial conditions" can be understood in many different ways. The Past Hypothesis understands it in a cosmological aspect: the entropy of the Universe around the Big Band was low. The Past Hypothesis goes back to Boltzmann; it has been forcefully defended by Albert and seen by Price "as the most important achievement of the late-twentieth-century physics". Earman, however, finds the hypothesis "not even false" since in most cosmological theories the notion of Boltzmann entropy cannot even be defined. An even bigger problem is that it is not at all clear how the initial low-entropy state of the Universe can account for the unidirectional increase of the entropy in *everyday* thermodynamic systems. An attempt to adjust this problem is Reichenbach's "branch system approach" according to which thermodynamic systems branch off from and return back into the rest of the Universe "inheriting" the entropy increase from it.

25. The Recurrence objection is based on Poincaré's Recurrence theorem (1889) stating that in a Hamiltonian system enclosed in a finite volume with a finite energy almost all states will eventually return to arbitrary close to itself.

Poincaré's Recurrence theorem. Let (Γ, μ, ϕ_t) a dynamical system with $\mu(\Gamma) < \infty$. Let $A \in \mathcal{F}$ be any measurable subset of Γ, and define the set B for a given time τ as

$$B := \{x \in A \,|\, \forall t > \tau : \phi_t(x) \notin A\}$$

Then $\mu(B) = 0$.

Proof. Denote $\phi_t(B)$ by B_t. Since no point returns from B back to A and hence to B, therefore there exists a large enough t such that $\mu(B \cap B_t) = 0$. It also follows that for any $n \in \mathbb{N}$: $\mu(B \cap B_{nt}) = 0$. Now, since ϕ_t is a bijection it also follows that for any two $n_1 < n_2 \in \mathbb{N}$: $\mu(B_{n_1 t} \cap B_{n_2 t}) = 0$, otherwise $\mu(B \cap B_{(n_2 - n_1)t})$ were not zero. But from Liouville's theorem $\mu(B_{nt}) = \mu(B)$ for any n. Since $\mu(\Gamma) < \infty$, therefore $\mu(B) = 0$. ■

Zermelo's Recurrence objection is making use of Poincaré's Recurrence theorem as follows. After a Poincaré cycle a finite system gets back 'very' close to its initial state. If the entropy $S_B(M(x))$ is continuous in x (assumption!), then it cannot increase monotonically. As Poincaré puts it:

> "According to this theory, to see heat pass from a cold body into a warm one, it will not be necessary to have the acute vision, the intelligence and the dexterity of Maxwell's demon; it will suffice to have a little patience."

Boltzmann responded by a rough calculation pointing out that "little patience" $= 10^{10^{19}}$ seconds for a cubic centimeter of air (microstates are defined here within 10 Å spatial position for each molecule and within 1-3% speed value).

Notes and further readings

For the objections to the Boltzmann program see Sklar (1993), Uffink (2007) and Frigg (2008, 117-119). For the Past Hypothesis see Albert (2000, Ch. 4); for a criticism see Earman (2006).

8 The Gibbs approach

26. Ensembles. While the Boltzmann approach considers one individual system, the Gibbs program start its analysis with an *ensemble* of systems. As Gibb (1902) writes:

> ”We may imagine a great number of systems of the same nature, but differing in the configurations and velocities which they have at a given instant, and differing not merely infinitesimally, but it may be so as to embrace every conceivable combination of configuration and velocities. And here we may set the problem, not to follow a particular system through its succession of configurations, but to determine how the whole number of systems will be distributed among the various conceivable configurations and velocities at any required time, when the distribution has been given for some one time”

What are these ensembles? They are "mental copies of the one system under consideration" (Schrödinger); they do not interact with each other.

The formalism of the Gibbs approach starts with a probability density function $\rho_t(x)$ on the $6n$ dimensional Γ-space (the Gibbs approach does not use the μ-space), where $x = (q, p)$. The essence of the Gibbs approach is that the observable quantities are associated to ensemble average of phase functions $f(x, t)$

$$\langle f \rangle_t = \int_\Gamma f \rho_t \, d\mu$$

Now, due to the Liouville's theorem

$$\frac{d\rho_t}{dt} = \frac{\partial \rho_t}{\partial t} + \{\rho_t, H\} = 0$$

(proved by the $6n$ dimensional continuity equation) ρ 'flows' as an incompressible fluid.[4] So the dynamics $\rho_t(x) := \rho_0(\phi_{-t}x)$ can be expressed in the Hamiltonian mechanics as

$$\frac{\partial \rho_t}{\partial t} = \{H, \rho\}$$

A probability density $\rho_t(x)$ is *stationary* if for all t and $x : \rho_0(\phi_{-t}x) = \rho_0(x)$ which means that

$$\frac{\partial \rho_t}{\partial t} = 0$$

Stationary probability densities characterize *equilibrium* since they yield constant averages:

$$\frac{\partial \langle f \rangle_t}{\partial t} = \int_\Gamma f \frac{\partial \rho_t}{\partial t} d\mu = 0$$

---

[4] The Poisson bracket $\{\rho, H\}$ is defined as $\{\rho, H\} := \sum_{i=1}^{2rN} \left(\frac{\partial \rho}{\partial q_i} \frac{H}{\partial p_i} - \frac{\partial \rho}{\partial p_i} \frac{H}{\partial q_i} \right)$

if f does not depend explicitly on time. The special cases are:

$$\begin{aligned}
\rho_E(x) &= \frac{1}{\omega(E)}\delta(H(x)-E) \quad \text{(microcanonical: } E, N \text{ is fixed)}\\
\rho_T(x) &= \frac{1}{Z(T)}e^{\frac{-H(x)}{T}} \quad \text{(canonical: } N \text{ is fixed, } E \text{ can vary)}\\
\rho_{T,\alpha}(x,N) &= \frac{1}{N!Z(T,\alpha)}e^{\frac{-H(x)}{T+\alpha N}} \quad \text{(grand-canonical: } E, N \text{ can vary)}
\end{aligned}$$

27. The Gibbs entropy. The fined-grained Gibbs entropy is defined as follows:

$$S_{G,f}(\rho) := -k\int_\Gamma \rho\log(\rho)\, d\mu$$

For the above constraints the appropriate distribution has the maximal fine-grained Gibbs entropy.

Now, how $S_{G,f}$ changes with the dynamics ϕ_t? Due to Liouville's theorem the fine-grained Gibbs entropy remains constant:

$$\frac{dS_{G,f}}{dt} = \frac{\partial S_{G,f}}{\partial t} = 0$$

(See (Zeh, 2007, 52)). So the fined-grained Gibbs entropy cannot tend towards equilibrium. Therefore Gibbs introduced his ink drop analogy: mixing two incompressible fluids: first take finite regions, then stir the fluids, then tend to zero with the volume of the regions ($\neq$ first tend to zero with the volume of the region, then stir the fluids!)

He introduced a *coarse-graining*:

$$\mathcal{CG} : \rho(x) \mapsto \hat{\rho}(x) = \sum_i \hat{\rho}(i)\mathbf{1}_{\omega_i}(x), \quad \text{where } \hat{\rho}(i) := \frac{1}{\delta\omega}\int_{\omega_i}\rho(x)dx$$

and by this the *coarse-grained Gibbs entropy*:

$$S_{G,c}(\rho) := S_{G,f}(\hat{\rho}) = -k\int_\Gamma \hat{\rho}\log(\hat{\rho})\, d\mu$$

Now, $S_{G,c}[\rho]$ is not constrained by Liouville's theorem! But does the coarse-grained Gibbs entropy evolve towards the equilibrium state? It is true that: $S_{G,c}[\rho] \geqslant S_{G,f}[\rho]$. (The equality holds only if the fine-grained distribution is uniform over the cells.) But $S_{G,c}[\rho]$ is non-decreasing only if the system is mixing.

Notes and further readings

Gibbs' seminal work is (Gibbs 1902).

9 Entropy and information

28. Entropy is a term defined in a whole bunch of disciplines from information theory to physics, biology and ecology. First we briefly list some notions of entropy used in SM.

1. Thermodynamic entropy is a macrovariable in thermodynamics defined for a given equilibrium state A of temperature T as:

$$S_T(A) := \int_O^A \frac{dQ}{T}$$

where O is a reference equilibrium state. The Planck version of the Second Law then reads as follows: For any process from state A to state B:

$$S_T(B) \geqslant S_T(A)$$

2. Fine-grained Boltzmann entropy is defined as the logarithm of the measure of that region of the phase space which is compatible with a given set of macrostates:

$$S_{B,f}(M_i(x)) := k \log \big(\mu_E(\Gamma_{M_i(x)})\big)$$

The subindex in μ_E is to account for the constraints such as the energy hypersurface. The logarithm is to make the entropy an extensive quantity. A microstate x has entropy by means of the macrostate M_i supervening on x.

3. The coarse-grained Boltzmann entropy was introduced by Boltzmann in his 1872 paper as a reaction to objections to his former works. For a given coarse-graining of the phase space any (fine-grained) microstate x determines an arrangement (a coarse-grained microstate) which further determines a state distribution. The coarse-grained Boltzmann entropy is the logarithm of that phase space region which is compatible with a given distribution $Z_{\{n_i\}}$:

$$S_{B,c}(Z_{\{n_i\}}(x)) := k \log \Big(\mu(\Gamma_{Z_{\{n_i\}}(x)})\Big)$$

In other words, the coarse-grained Boltzmann entropy is the number of arrangements compatible with a given distribution *times* the volume of the elementary cell. Again, a microstate x has entropy by means of the distribution $Z_{\{n_i\}}(x)$ as a function on Γ.

The coarse-grained Boltzmann entropy is a special case of the fine-grained Boltzmann entropy where macrostates are given by state distributions (partition functions).

4. The fine-grained Gibbs entropy refers to an ensemble rather than to an individual system and is defined as:

$$S_{G,f}(\rho(x)) := -k \int_\Gamma \rho(x) \log(\rho(x))\, d\mu$$

where $\rho(x)$ is the probability density. The fine-grained Gibbs entropy can be related to an individual system if $\rho(x)$ can be interpreted either as the frequency of finding a given individual system in the microstate x within the whole ensemble; or as the fraction of time a given system spends in x. Both interpretations are dubious. Moreover, due to the Liouville theorem the fine-grained Gibbs entropy is constant over time.

5. The coarse-grained Gibbs entropy is defined as the fine-grained Gibbs entropy of the coarse-grained probability density $\hat{\rho}$:

$$S_{G,c}(\rho) := S_{G,f}(\hat{\rho}) = -k \int_{\Gamma} \hat{\rho} \log(\hat{\rho})\, d\mu$$

It is introduced to tackle the problem of the constancy of the fine-grained Gibbs entropy.

29. Information theory also uses various notions of entropy such as the Shannon entropy, Rényi entropy, Hartley entropy, von Neumann entropy, etc. Let us see some of them.

6. The Shannon entropy is devised to answer the question as to how much information is contained in a set of messages emitted by a source.

First, we need to define the information we obtain by observing an event having probability p. This can be settled by assuming that the information as a function of p satisfies the following properties:

1. *Positivity.* $I(p) \geqslant 0$.
2. *Normalization.* $I(1) = 0$.
3. *Additivity.* The information adds up for independent events: $I(p_1 \cdot p_2) = I(p_1) + I(p_2)$.
4. *Continuity.* $I(p)$ is a monotonic and continuous function of p.

Then one can show that the information has the following form:

$$I(p) = \log_b (1/p) = -\log_b(p)$$

Due to

$$\log_a(x) = \log_a(b) \log_b(x) = c \log_b(x)$$

the different bases simply measure the information in different units. Hence we drop the basis.

Next suppose that a source produces n different messages (events) with a probability distribution $p = (p_1, p_2 \dots p_n)$. The Shannon entropy is the average amount of information we obtain from the source:

$$S_{Sh}(p) := -\sum_{i=1}^{n} p_i \log p_i$$

The Shannon entropy can be generalized for the continuous case as

$$S_{Sh}(p) := -\int_{\mathbb{R}} \rho(x)\log(\rho(x))\,dx$$

Now, the discrete Shannon entropy is related to the coarse-grained Boltzmann entropy and the continuous Shannon entropy is related to the fine-grained Gibbs entropy (modulo the Boltzmann constant k). The second is obvious; to see the first suppose that n is large, hence by means of the Stirling formula the coarse-grained Boltzmann entropy can be written as:

$$S_{B,c}\left(Z_{\{n_i\}}\right) = \sum_i n_i \log n_i + c \tag{17}$$

Introducing the probability $p_i = \frac{n_i}{n}$ of finding a randomly chosen particle in cell ω_i we obtain:

$$S_{B,c}\left(Z_{\{n_i\}}\right) = \sum_i p_i \log p_i + c' \tag{18}$$

which is the discrete Shannon entropy.

Before going over to the next type of entropy it is instructive to introduce a notion called *majorization* which measures how mixed a given probability distribution is. Let p and q be two discrete probability distributions on a set of n events. We say that q majorizes p, $p \preceq q$ iff

$$\sum_{i=1}^{k} p_i^{\downarrow} \;\leqslant\; \sum_{i=1}^{k} q_i^{\downarrow} \qquad \forall k = 1 \ldots n$$

where $\downarrow$ in the superscript means decreasing order. It is easy to see that majorization is a partial order on the set of probability distributions up to permutations and the smallest and greatest elements are the *uniform distribution* ($p_i = 1/n$)and the *pure state* ($p = p^2$), respectively:

$$(1/n, 1/n, \ldots, 1/n) \preceq p \preceq (1, 0, \ldots, 0)$$

Now, one is looking for real-valued functions f on the probability distributions compatible with majorization in the sense that

$$p \preceq q \quad \Longleftrightarrow \quad f(p) \geqslant f(q) \tag{19}$$

In other words, f should be greater for more mixed probability distributions. Functions satisfying (19) are called *Schur-concave.* Schur-concavity is the definitive property of the general notion of entropy. Hence f represents mixedness in the same sense as utility represents preference order. But note that entropy can be compared for any pair of probability distributions not just for those can be related by majorization. Majorization is a partial ordering, whereas entropy totally orders the distributions.

7. Rényi entropy is a generalized entropy in the above sense:

$$S_R(p) := \frac{1}{1-q} \log \sum_i p_i^q \qquad q > 0$$

For $q \to 1$ it gives back the Shannon entropy and for $q \to 0^+$ if gives back the *Hartley entropy*

$$S_H(p) := \log \big| \{ i \,|\, p_i \neq 0 \} \big|$$

8. Von Neumann entropy is the generalization of Shannon entropy for the quantum case. In QM probability distributions are replaced by states represented by density matrices. For a given state ρ the von Neumann entropy is defined as follows:

$$S_{vN}(\rho) := -Tr(\rho \log \rho)$$

The quantum analogue of the other classical entropy concepts can be defined in a similar way.

Notes and further readings

This section is partly based on (Frigg and Werndl, 2011) and partly on (Bengtsson and Życzkowski, 2006).

Appendix A: The Ehrenfests' dog-flea model

30. The Ehrenfests' dog-flea model (also known as *urn model*) provides an instructive model for a process approaching to equilibrium. Two dog, a white (W) and a black (B), share a population of fleas indexed from 1 to N. Suppose that N is even. At each second a number from 1 to N is chosen randomly and the flea labeled by this number jumps on the other dog. The *microstates* of the model are the specification of which flea is on which dog; and the *macrostates* are the numbers (W, B) of fleas on the two dogs. Since $B = N - W$, it is enough to consider W as a macrostate. There are 2^N microstates and $N + 1$ macrostates of the system.

First note that the process defined in the model is stochastic, moreover Markovian with time independent transition probabilities:

$$p\big((W-1)_{t+1}\,|\,W_t\big) \;=\; \frac{W}{N} \tag{20}$$

$$p\big((W+1)_{t+1}\,|\,W_t\big) \;=\; 1-\frac{W}{N} \tag{21}$$

One can see that the transition probabilities drive any macrostate W in the direction of the macrostate $\frac{N}{2}$ which is called the *equilibrium macrostate.*

Given the probability distribution $\big\{p\big(W_t\big)\big\}$ of the macrostates at time t one can calculate by means of (20)-(21) the probability distribution at time $t + 1$:

$$\begin{aligned} p\big(W_{t+1}\big) &= p\big(W_{t+1}\,|\,(W-1)_t\big)\, p\big((W-1)_t\big) + p\big(W_{t+1}\,|\,(W+1)_t\big)\, p\big((W+1)_t\big) \\ &= \left(1-\frac{W-1}{N}\right) p\big((W-1)_t\big) + \left(\frac{W+1}{N}\right) p\big((W+1)_t\big) \end{aligned} \tag{22}$$

Let

$$\langle W\rangle_t := \sum_{W=0}^{N} p\big(W_t\big)\, W \tag{23}$$

denote the expectation value of the macrostate W at time t. Using (20)-(21) and (22) one can calculate the evolution of the expectation value:

$$\begin{aligned} \langle W\rangle_{t+1} &= \sum_{W=0}^{N} p\big(W_{t+1}\big)\, W \\ &= \sum_{W=0}^{N} \left[\left(1-\frac{W-1}{N}\right) p\big((W-1)_t\big) + \left(\frac{W+1}{N}\right) p\big((W+1)_t\big)\right] W \\ &= \sum_{W=0}^{N} \left[\left(1-\frac{2}{N}\right) W + 1\right] p\big(W_t\big) \\ &= \left(1-\frac{2}{N}\right) \langle W\rangle_t + 1 \end{aligned} \tag{24}$$

where in the shifting of the summation index in the third row we assumed that the macrostates $W = -1$ and $W = N+1$ are not admissible. The solution of the difference equation (24) is:

$$\langle W \rangle_t = \left(1 - \frac{2}{N}\right)^t \left(\langle W \rangle_{t=0} - \frac{N}{2}\right) + \frac{N}{2} \tag{25}$$

showing that starting from any macrostate the expectation is approaching to the equilibrium macrostate $W = \frac{N}{2}$.

One also can calculate the unique time-invariant probability distribution with respect to the transition probabilities (20)-(21) which is:

$$p\big(W\big) = \frac{1}{2^N}\binom{N}{W} \tag{26}$$

since using (22) one obtains:

$$\begin{aligned} p\big(W_{t+1}\big) &= p\big(W_{t+1} \,|\, (W-1)_t\big)\, p\big((W-1)_t\big) + p\big(W_{t+1} \,|\, (W+1)_t\big)\, p\big((W+1)_t\big) \\ &= \left(1 - \frac{W-1}{N}\right)\binom{N}{W-1} + \left(\frac{W+1}{N}\right)\binom{N}{W+1} \\ &= \frac{1}{2^N}\binom{N}{W} = p\big(W_t\big) \end{aligned} \tag{27}$$

Using the time invariant measure (26) and the Bayes rule the backward transition probabilities turn out to be equal to the forward transition probabilities:

$$\begin{aligned} p\big((W-1)_{t-1} \,|\, W_t\big) &= p\big(W_t \,|\, (W-1)_{t-1}\big)\, \frac{p\big((W-1)_{t-1}\big)}{p\big(W_t\big)} \\ &= \left(1 - \frac{W-1}{N}\right) \frac{\binom{N}{W-1}}{\binom{N}{W}} = \frac{W}{N} = p\big((W-1)_{t+1} \,|\, W_t\big) \\ p\big((W+1)_{t-1} \,|\, W_t\big) &= p\big(W_t \,|\, (W+1)_{t-1}\big)\, \frac{p\big((W+1)_{t-1}\big)}{p\big(W_t\big)} \\ &= \frac{W+1}{N} \frac{\binom{N}{W+1}}{\binom{N}{W}} = 1 - \frac{W}{N} = p\big((W+1)_{t+1} \,|\, W_t\big) \end{aligned}$$

That is the process is *time-reversal invariant.*

Now, by means of the invariant measure (26) we can calculate the probability for a macrostate W_t of having come from $(W-1)_{t-1}$ and returning to $(W-1)_{t+1}$:

$$\begin{aligned} p\big((W-1)_{t+1} \wedge (W-1)_{t-1} \,|\, W_t\big) &= \frac{p\big((W-1)_{t+1} \wedge W_t \wedge (W-1)_{t-1}\big)}{p\big(W_t\big)} \\ &= \frac{p\big((W-1)_{t+1} \,|\, W_t\big)\, p\big(W_t \,|\, (W-1)_{t-1}\big)\, p\big((W-1)_{t-1}\big)}{p\big(W_t\big)} \\ &= \frac{\frac{W}{N}\,(1-\frac{W-1}{N})\,\frac{1}{2^N}\binom{N}{W-1}}{\frac{1}{2^N}\binom{N}{W}} = \frac{W^2}{N^2} \qquad (28) \end{aligned}$$

Similarly,

$$p\big((W-1)_{t+1} \wedge (W+1)_{t-1} \,|\, W_t\big) = \frac{W(N-W)}{N^2} \qquad (29)$$

$$p\big((W+1)_{t+1} \wedge (W-1)_{t-1} \,|\, W_t\big) = \frac{W(N-W)}{N^2} \qquad (30)$$

$$p\big((W+1)_{t+1} \wedge (W+1)_{t-1} \,|\, W_t\big) = \frac{(N-W)^2}{N^2} \qquad (31)$$

For $W_t > \frac{N}{2}$, (28) is bigger than (31); for $W_t < \frac{N}{2}$, (31) is bigger than (28). This means that for any macrostate W_t the most probable history, especially if it is far away from equilibrium, is to having arrive and returning to a macrostate which is closer to the equilibrium macrostate. A non-equilibrium macrostate is most probably an endpoint of a fluctuation.

The dog-flea model, though stochastic, is an *ergodic* Markov process in the sense it is possible to get from every macrostate to every other macrostate with positive probability.

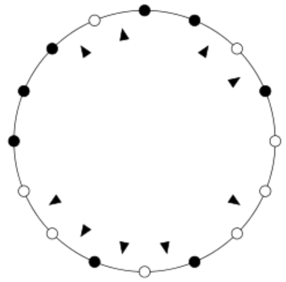
Notes and further readings

The dog-flea model was invented by the Ehrenfests (1907). For a nice treatment and computer simulation see (Emch and Liu, 2002).

Appendix B: The Kac ring model

31. The Kac ring model is a non-ergodic, explicitly solvable model with a deterministic and time-reversal invariant dynamics displaying irreversible thermodynamic behavior. It was introduced by Marc Katz in 1956.

Consider a ring with N equidistant balls on it each colored either white or black. Suppose that $M < N$ of the edges connecting the balls contains a *marker*. In each second the ring revolves one step counterclockwise taking each ball to the nearest site. When a ball passes a marker, it changes color. (If there were only one marker randomly relocated on the ring at each step, we would get back the dog-flea model.)



The dynamics of the Kac ring is time-reversal invariant and has a Poincaré cycle of length $2N$ since after $2N$ steps each ball returns to its initial state and passes each marker twice.

Let W_t and B_t denote the number of white and black balls at time t, respectively. Let the macrostate of the system be the difference

$$G_t := W_t - B_t \tag{32}$$

of the number of the white and black balls ("greyness"). The system – similarly to the dog-flea model – has 2^N *microstates* and $N + 1$ *macrostates*.

Let us see how the macrostate of the system evolves with time. Denote by W_t^* and B_t^* the number of white and black balls in front of a marker at time t. Then the number of the white and black balls changes from t to $t + 1$ as follows:

$$\begin{aligned} W_{t+1} &= W_t - W_t^* + B_t^* && (33)\\ B_{t+1} &= B_t - B_t^* + W_t^* && (34) \end{aligned}$$

and hence the macrostate evolves as

$$G_{t+1} = W_{t+1} - B_{t+1} = G_t + 2W_t^* - 2B_t^* \tag{35}$$

32. The Boltzmann approach. To solve (35) one can use two approaches called the *Boltzmann approach* (master equation approach) and the *Gibbs approach*. Let is start with the Boltzmann approach.

To solve (35) one should know W_t^* and B_t^* at each t. Instead, however, let us make the following assumption: *the markers are randomly distributed on the ring.* That is the number of a white balls being in front of a marker is just the density of the markers on the ring times the number of the white ball; and similarly for the black balls:

$$\frac{M}{N} = \frac{W_t^*}{W_t} = \frac{B_t^*}{B_t} \tag{36}$$

This assumption is analogue to Boltzmann's famous *Stosszahlanzatz* or *molecular chaos* assumption. It states that the distribution of the markers across the balls of different color at time t is independent of what happened to the balls before t. It is as if the markers were randomly reshuffled on the ring before every step; as if we have N dog-flea models at each step. Obviously (36) cannot hold at each t but the hope is that it is a typical behavior for large N.

Now, by means of (36) the equation (35) becomes

$$G_{t+1} = \left(1 - 2\frac{M}{N}\right) G_t \tag{37}$$

leading to the solution

$$G_t = \left(1 - 2\frac{M}{N}\right)^t G_0 \tag{38}$$

As t tends to infinity, G_t tends to zero showing that the number of the white and black balls becomes equal on the long run.

But this cannot be right: At $t = 2N$ the initial configuration returns! A case for Loschmidt's Reversibility objection.

33. The Gibbs approach. Let us see the process at the microlevel. First let us extend the phase space to include also the different distributions of the markers on the rings and try to recover (38) as a *statistical* equation arising from a suitable probability distribution on the distributions of the markers on the rings.

First, introduce the following two variables, one for the color of a given ball:

$$C_{i,t} \;=\; \begin{cases} 1 & \text{if the ball at site } i \text{ at time } t \text{ is white} \\ -1 & \text{if the ball at site } i \text{ at time } t \text{ is black} \end{cases} \tag{39}$$

and one for whether there is a marker on a given edge of the ring:

$$M_i \;=\; \begin{cases} 1 & \text{if there is a marker at the edge connecting site } i \text{ and } i+1 \\ -1 & \text{if there is no marker at the edge connecting site } i \text{ and } i+1 \end{cases} \tag{40}$$

Now, the one-step evolution of the system is

$$C_{i+1,\,t+1} = M_i\, C_{i,t} \tag{41}$$

which leads to the following evolution of a given ball:

$$C_{i,t} = M_{i-1}\, M_{i-2} \dots M_{i-t}\, C_{i-t,0} \tag{42}$$

We obtain the macrostate G_t by adding up the balls:

$$G_t = \sum_{i=1}^{N} C_{i,t} = \sum_{i=1}^{N} M_{i-1}\, M_{i-2} \dots M_{i-t}\, C_{i-t,0} \tag{43}$$

Obviously, $G_{2N} = G_0$.

Now, consider $\langle G \rangle_t$, the average of the difference G at time t over the different distributions of markers on the ring for a given initial configuration $\{C_{i,0}\}$ of the balls. $\langle G \rangle_t$ can be written as follows:

$$\langle G \rangle_t = \sum_{i=1}^{N} \langle C_{i,t} \rangle = \sum_{i=1}^{N} \langle M_{i-1}\, M_{i-2} \dots M_{i-t} \rangle\, C_{i-t,0} = \langle M_{i-1}\, M_{i-2} \dots M_{i-t} \rangle\, G_0 \tag{44}$$

Suppose that the M markers are randomly placed on the N edges; that is they follow a binomial distribution. Hence for $t < N$ consecutive edges we have:

$$\langle M_{i-1}\, M_{i-2} \dots M_{i-t} \rangle = \sum_{j=1}^{t} (-1)^j \binom{t}{j} \left(\frac{M}{N}\right)^j \left(1 - \frac{M}{N}\right)^{t-j} = \left(1 - 2\frac{M}{N}\right)^t \tag{45}$$

by which we get the statistical version of (38):

$$\langle G \rangle_t = \left(1 - 2\frac{M}{N}\right)^t \langle G \rangle_0 \tag{46}$$

Therefore the *Stosszahlanzatz* (36) can be related to an ensemble over distributions of markers on the ring.

In what sense does the statistical equation (46) justify equation (38)? In the sense of *typicality*. That is if the majority of the evolution of the individual systems remains close to the average behavior of the ensemble. In other words, if the average is peaked. However, the variance of the ensemble scales like N for the Kac ring, so the standard deviation like $\sqrt{N}$ as long as t is smaller than the half-recurrence time. This means that for short time and large N, the average behaviour is typical.

The coarse-grained Boltzmann entropy of the Kac ring can be defined via the *partition function* $Z_G := \binom{N}{W}$ as

$$S_{B,c}\,(Z_G) = \log(Z_G) = -N[p \log p + (1-p)\log(1-p)]$$

where $p := \frac{W}{N}$ and we used the Stirling formula. It is maximal for $p = \frac{1}{2}$. One can see that the coarse-grained Boltzmann entropy is extensive (linear in N).

Finally, let us compare the Boltzmann approach and the Liouville approach. In the Boltzmann approach we (i) first take a system; (ii) apply the coarse-graining to get the macrostates; (iii) let the system evolve macrodynamically; (iv) finally force independence of the subsystems in each step by demanding the *Stosszahlanzatz*. In the Liouville approach we (i) first take an ensemble of the systems; (ii) let them evolve microdynamically; (iii) apply the coarse-graining to each to get the macrostates; and (iv) finally calculate the statistical averages. If the variance is small, the process can be regarded typical.

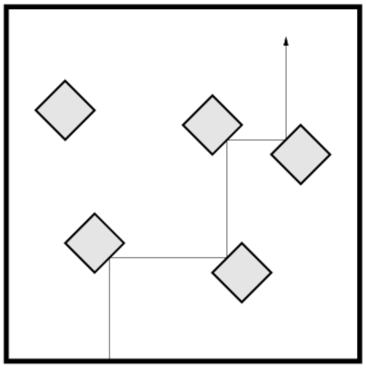
So in the two approaches coarse-graining and time evolution are swapped. Generally, however, they do not commute!

Notes and further readings

A beautiful treatment of the Kac ring model can be found in (Gottwald and Oliver, 2009). Kac' original paper is (Kac, 1956). See also (Brickmont, 1995).

Appendix C: The wind-tree model

34. The wind-tree model is again a deterministic model introduced by the Ehrenfests (1902). The model consists of a two dimensional gas with N particles ("wind") and M immobile scatterers ("trees") in a container of area A. The particles can move in four different directions: East (1), North (2), West (3) and South (4). The scatterers are square-shaped; their sides are of length a and make $45°$ with the above directions. Both the particles and the scatterers are randomly distributed. The particles do not interact with one another, but rebound from the scatterers elastically. By the collision the scatterers remain unaffected.



Let $f_i(t)$ be the number of particles moving in direction $i = 1, 2, 3, 4$ at time t. $\sum_i f_i(t) = N$. Let $N_{ij}(\Delta t)$ denote the number of particles that due to a collision are scattered from direction i to a neighboring direction $j = i \pm 1 \pmod 4$ between time t and $t + \Delta t$. Any such particle is in a parallelogram with base $v\Delta t$ and height $a/\sqrt{2}$ attached to the appropriate side of a scatterer. How many are they?

Suppose that the total area of the scatterers is very small compared to the total area of the container, $Ma^2 \ll A$. Again we apply an independence assumption analogous to the *Stosszahlanzatz*:

$$\frac{N_{ij}(\Delta t)}{f_i(t)} = \frac{M \frac{a}{\sqrt{2}} v \Delta t}{A}$$

That is at any moment the ratio of those particles that will be scattered from direction i to direction j within the time interval Δt relative to all particles traveling in direction i is just the ratio of the total area of all the appropriate parallelograms to the area of the container. From this we get that

$$N_{ij}(\Delta t) = \left(\frac{M}{A}\frac{a}{\sqrt{2}}v\right) f_i(t)\Delta t =: k f_i(t)\Delta t$$

Hence, the balance equation reads as follows:

$$\begin{aligned} f_i(t+\Delta t) &= f_i(t) + N_{i,i+1}(\Delta t) + N_{i,i-1}(\Delta t) - N_{i+1,i}(\Delta t) - N_{i-1,i}(\Delta t) \\ &= f_i(t) + k\big(f_{i+1}(t) + f_{i-1}(t) - 2f_i(t)\big)\Delta t \end{aligned}$$

It can be shown that the H-function

$$H := \sum_i f_i \log f_i$$

is monotonically decreasing until it reaches the minimum value at $f_i = \frac{N}{4}$.

Notes and further readings

The wind-tree model was also invented by the Ehrenfests (1907). See for example (Brown, Myrword and Uffink, 2009).

Appendix D: Questions

Pick one of the following questions and write a 1000 word discussion paper on it.

Q1. Where does coarse-graining come from? Limited capacity of the observer? Does equilibrium make sense to creatures with unlimited capacity? Does coarse-graining have any physical basis or it is plainly arbitrary?

Bibliography

D. Albert, *Time and Chance* (Harvard University Press, 2000).

I. Bengtsson and K. Życzkowski, *Geometry of quantum states*, (Cambridge: Cambridge University Press, 2006).

R. Frigg and C. Werndl, “Entropy: a guide for the perplexed,” in C. Beisbart, S. Hartmann, (eds.) *Probabilities in physics* (Oxford: Oxford University Press, 2011).

J. Brickmont, “Science of Chaos or Chaos in Science?” *Physicalia Magazine*, **17**, 159-208, 1995.

H. R. Brown, W. Myrvold, and J. Uffink, “Boltzmann’s H-theorem,its discontents, and the birth of statistical mechanics,” *Stud. Hist. Phil. Mod. Phys.*, **40**, 174-191 (2009).

J. Butterfield, *Philosophy of Thermal Physics*, (lecture notes).

G. G. Emch and C. Liu, *The logic of Thermostatistical Physics*, (Springer-Verlag Berlin Heidelberg 2002).

P. Ehrenfest and T. Ehrenfest, “Über zwei bekannte Einwände gegen das Boltzmannsche H-Theorem,” *Physikalishce Zeitschrift*, **8**, 311-314 (1907).

R. Frigg, A Field Guide to Recent Work on the Foundations of Statistical Mechanics, in: Rickles, Dean (ed.): *The Ashgate companion to contemporary philosophy of physics.* (Ashgate, 2008, 99-196).

J. W. Gibbs, (1902) *Elementary Principles in Statistical Mechanics*, (Woodbridge: Ox Bow Press, 1981).

G. A. Gottwald and M. Oliver, ”Boltzmann’s dilemma – an introduction to statistical mechanics via the Kac ring,” *SIAM Rev.*, **51(3)**, 613–635 (2009).

M. Kac, ”Some remarks on the use of probability in classical statistical mechanics,” *Acad. Roy. Belg. Bull. Cl. Sci. (5) 42* 356–361 (1956).

K. Simonyi, *A Cultural History of Physics* (CRC Press Taylor & Francis Group, 2012).

L. Sklar, *Physics and Chance, Philosophycal Issues in the Foundations of Statistical Mechanics* (Cambridge University Press, 1993).

J. Uffink, ”Compendium of the Foundations of Classical Statistical Physics,” in: Jeremy Butterfield and John Earman (eds.): *Philosophy of Physics* (Elsevier, 2007).

J. Uffink, ”Bluff your way with the Second Law of Thermodynamics,” *Stud. Hist. Phil. Mod. Phys.*, **32**, 305-394.

H. D. Zeh, *The Physical Basis of the Direction of Time* (Springer, 2007).