

Introduction to the Philosophy of Thermodynamics and Statistical Physics

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)

Course description

For the speaker:

- Read the chosen chapter one week before class and come up with three questions which help others in reading the text (For example: *Which are the competing historical approaches towards the nature of heat and what are the arguments for each?*). Make a pdf file with your name in the text and send it to me (gsz@szig.hu) one week before class. I will post them on the web page of the seminar (<http://hps.elte.hu/gszabo/PhysicsandChance.html>)
- Disseminate a (maximum two page) handout before your talk.
- Talk no less than 20 and no more than 30 minutes.
- You need not cover all of the material. Pick some points which you find interesting.
- Chair the discussion in class.

For everyone, for every week:

- Pick one question and write a 500 word answer to it. (If you find all questions unilluminating, write simply a 500 word summary of the chapter.) Send it to me (gsz@szig.hu) not later than one day before class (that is until Sunday 9.30).
- How to read philosophical texts? *The point of intractability.*

”What should you seek when you start reading in a new topic in philosophy of science? The emphasis should be on identifying the central theses and arguments. But how deeply should you read? I have found one rule of thumb very helpful. In any field, there are easy and obvious results. They are typically picked up and published early. The sign of maturity of a field is that none of them are left. Rather one develops a sense of a deep intractability that blocks further progress. You should seek to read to this point of intractability and try to find how that intractability arises. It is generally manifested in a ”damned if you do and damned if you don’t” dynamic. There will be a proliferation of different viewpoints, each designed to circumvent the difficulty, but with none commanding universal assent.

The classic philosophical problem, Hume’s problem of induction, illustrates this. How can inductive inference be justified? Any purported justification must call upon other means of inductive justification, so that they are circular or trigger an infinite regress. Or, if you accept that no justification is admissible, then it would seem we have no reason to believe inductive inference. So can you justify induction? You are damned if you try; and damned if you don’t.

The common experience in entering a new field in philosophy of science is that you see lots of easy results. Exactly because they are easy, chances are that they are widely known. When you find that progress has been stalled by apparently intractable problems, rejoice! You have come to the point where novel contributions are possible.” (John Norton)

1 Classical thermodynamics

- Birth: industrial revolution (late 18th, early 19th century) (*cf. BBC, In our Time: Heat*)
- 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)
- Heroes: Carnot, Clausius, Kelvin, Planck, Gibbs, Carathéodory, Lieb and Yngvason

1.1 Basic concepts

- 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: $pV = nRT = NkT$
- Equilibrium: thermodynamic coordinates are constant over time
- Quasi-static process: the system is in equilibrium throughout the process ('Moving a cup of tea filled to the brim without spilling')
- Reversible process: the process $\langle s_i, Z_i \rangle \xrightarrow{\mathcal{P}} \langle s_f, Z_f \rangle$ is *irreversible* if there exists another process \mathcal{P}' such that $\langle s_f, Z_f \rangle \xrightarrow{\mathcal{P}'} \langle s_i, Z_i \rangle$ complete recovering the initial state of the system *and the environment* (frictionless piston)
 - Reversible \implies quasi-static; but quasi-static $\not\Rightarrow$ reversible (piston subject to friction)
 - A reversible process gives the maximum limit of work that a system can produce
 - A reversible process need not be adiabatic
 - Reversible \neq time reversal invariant

1.2 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)
 - Temperature: choose a system with a convenient property x (length) and take (empirical) temperature θ to be a linear function of x : $\theta(x) = ax + b$
 - Ideal gas temperature: T is defined through $pV = nRT$ and $T_s - T_i = 100$
- 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!)

1.3 Carnot cycle and entropy

- Carnot cycle
 - Efficiency of heat engine: $\eta = \frac{W}{Q_2} = 1 - \frac{Q_1}{Q_2} \leq 1$
 - Coefficiency of refrigerator: $\kappa = \frac{Q_1}{W} = \frac{Q_1}{Q_2 - Q_1} = \frac{1}{\eta} - 1 \geq 0$
- Carnot Principle: Let the temperature of two heat reservoirs be T_1 and T_2 (with $T_1 < T_2$). Then:
 - (i) all heat engines operating in a reversible cycle have the same efficiency depending only on the temperatures T_1 and T_2 ;
 - (ii) all other heat engines have an efficiency which is less than or equal to that obtained by reversible cycles.
- The Carnot Principle is the consequence of the Second Law
- Thermodynamic temperature: (the efficiency of the reversible Carnot cycle depends only on the temperature) $\frac{T_s - T_i}{T_s} = 1 + \frac{Q_1}{Q_2} = 0.268$ and $T_s - T_i = 100$ (identical with the ideal gas temperature)
- 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: $\sum_i \frac{\pm Q_i}{T_i} \longrightarrow \oint \frac{dQ}{T}$
 - 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: $S(B) - S(A) \geq \int_A^B \frac{dQ}{T}$
 - The Second Law is exceptionless
 - Entropy is a property of an individual system
 - Entropy does not fluctuate
- For adiabatic ($\Delta Q = 0$) processes: $\Delta S \geq 0$
 - Reversible adiabatic expansion: $\Delta Q = 0$ therefore $\Delta S = 0$
 - Heat conduction (irreversible process): $\Delta S = \Delta S_1 + \Delta S_2 = \frac{\Delta Q}{T_1} - \frac{\Delta Q}{T_2} = \Delta Q \left(\frac{1}{T_1} - \frac{1}{T_2} \right) > 0$
 - Free expansion (irreversible process): $\Delta S = nR \ln \frac{V_2}{V_1} > 0$

2 Kinetic theory

2.1 Basic concepts

- The aim of the kinetic theory: to give an account for the thermodynamical behaviour of macroscopic systems in terms of the dynamical laws governing their microscopic constituents and some probabilistic assumptions
- 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?
- Two ways of doing this:
 - Maxwell, Boltzmann: single system (μ -space) approach
 - Gibbs: ensemble (Γ -space) approach
- μ -space
 - μ -space: $\mu \ni (\mathbf{r}, \mathbf{v}) = (q_1, q_2, q_3; v_1, v_2, v_3)$, 6 dimensional
 - Velocity distribution, density function: $f(\mathbf{r}, \mathbf{v}, t)$
 - $f(\mathbf{r}, \mathbf{v}, t)$ is continuous and normalized
 - $d^3r d^3v$: volume element – 'large and small'
- Γ -space
 - Phase space: $\Gamma \ni x := (q_1, q_2 \dots q_n; p_1, p_2 \dots p_n)$, $6N$ dimensional
 - Dynamics: $\dot{q}_i = \frac{\partial H}{\partial p_i}$, $\dot{p}_i = -\frac{\partial H}{\partial q_i} \implies \phi_t(x)$: Hamiltonian flow
 - Constraints, energy hypersurface: $\Gamma_E := \{x \in \Gamma \mid H(x) = E\}$
 - Lebesgue (volume) measure on Γ : μ
 - Lebesgue measurable subsets of Γ : $A \in \mathcal{F}$
 - Liouville's theorem: The Lebesgue measure is invariant under the Hamiltonian flow: for any t and $A \in \mathcal{F}$: $\mu(\phi_t(A)) = \mu(A)$
 - Dynamical system: $(\Gamma, \mathcal{F}, \mu, \phi_t)$
 - Induced measure on Γ_E : $\mu_E(A_E) := \int_{A_E} \frac{d\sigma_E}{\|\nabla H\|}$ where $A_E \in \mathcal{F}_E^1$. Also μ_E is invariant: $\mu_E(\phi_t(A_E)) = \mu_E(A_E)$ for any t and $A_E \in \mathcal{F}_E$

3 Maxwell

3.1 Maxwell, 1860

- μ -space approach
- Until Maxwell: uniform velocity distribution
- Maxwell's question: What form $f(\mathbf{r}, \mathbf{v}, t) = f(\mathbf{r}, \mathbf{v})$ takes in equilibrium?
- Three "precarious assumptions":
 1. Homogeneity of space: $f(\mathbf{r}, \mathbf{v}) = f(\mathbf{v})$
 2. Isotropy of space: $f(\mathbf{v}) = g(v) = g\left(\sqrt{v_x^2 + v_y^2 + v_z^2}\right)$
 3. Independence of directions: $g(v) = h(v_x)h(v_y)h(v_z)$ (*)
- Solution of (*): $f(\mathbf{v}) = \left(\frac{m}{2\pi kT}\right)^{\frac{3}{2}} e^{-\frac{mv^2}{2kT}}$

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

- Remarks:
 - Maxwell adopts the frequency interpretation of probability $f(\mathbf{v})d^3v$, but Assumption 3 sits better with the classical interpretation
 - Collisions play no role in the derivation

3.2 Maxwell, 1867

- New model: point particles with equal masses interacting by means of a repulsive central force proportional to Δr^{-5}
- Number of collisions: $N(\mathbf{v}_1, \mathbf{v}_2) = N^2 f(\mathbf{v}_1) f(\mathbf{v}_2) |\mathbf{v}_2 - \mathbf{v}_1| d\mathbf{v}_1 d\mathbf{v}_2 dV dt$
- *Stosszahlansatz* (aka. molecular chaos) is used: the number of collisions during a time dt in spatial volume dV throwing molecules with initial velocities \mathbf{v}_1 and \mathbf{v}_2 within d^3v_1 and d^3v_2 into \mathbf{v}'_1 and \mathbf{v}'_2 within $d^3v'_1$ and $d^3v'_2$ is proportional to the product of the number of molecules with velocity \mathbf{v}_1 within d^3v_1 and those with velocity \mathbf{v}_2 within d^3v_2 and that spatial volume element
- For stationary distribution: $N(\mathbf{v}_1, \mathbf{v}_2) = N(\mathbf{v}'_1, \mathbf{v}'_2) \implies f(\mathbf{v}_1) f(\mathbf{v}_2) = f(\mathbf{v}'_1) f(\mathbf{v}'_2)$, for which (*) is a solution
- Remarks:
 - Collisions do play a role
 - The Maxwell distribution is not the unique stationary distribution
 - Boltzmann's counterexample: molecules lined up so that they only collide centrally and move perpendicularly between parallel walls. Then $f(\mathbf{v}) = \frac{1}{2}(\delta(\mathbf{v} - \mathbf{v}_0) + \delta(\mathbf{v} + \mathbf{v}_0))$
 - Maxwell was convinced that the derivation of the Second Law from mechanical principles was impossible:

”It is rare sport to see those learned Germans contending the priority of the discovery that the 2nd law of [thermodynamics] is the ‘Hamiltonsche Prinzip’, [...] The Hamiltonsche Prinzip, the while, soars along in a region unvexed by statistical considerations, while the German Icarus flap their waxen wings in nephelococcygia amid those cloudy forms which the ignorance and finitude of human science have invested with the incommurable attributes of the invisible Queen of Heaven”

”The Second Law has the same degree of truth as the statement that if you throw a tumblerful of water into the sea, you cannot get the same tumblerful of water out again.”
 - \implies Maxwell's demon (7.10)

4 Boltzmann

4.1 Boltzmann, 1872

- μ -space approach
- Boltzmann's program: whatever the initial velocity distribution of gas molecules, it approaches (due to collision) to the Maxwell-Boltzmann distribution (non-equilibrium theory)
- Boltzmann equation: $\frac{\partial f}{\partial t} + \mathbf{v}(\nabla_{\mathbf{v}} f) + \frac{\mathbf{F}}{m} \nabla f = \left(\frac{\partial f}{\partial t} \right)_{coll}$
- The collision term (using the *Stosszahlansatz*) is: $\left(\frac{\partial f}{\partial t} \right)_{coll} = N \int d\Omega \int d^3v_2 \sigma(\Omega) |\mathbf{v}_2 - \mathbf{v}_1| (f'_1 f'_2 - f_1 f_2)$

- $\sigma(\Omega)$: differential cross section (for hard spheres: $\frac{a^2}{4}$)
- $f_1 f_2$: the number of pairs of molecules with velocity \mathbf{v}_1 and \mathbf{v}_2 , respectively colliding in the volume d^3v at time t (outgoing molecules) (*Stosszahlansatz!*)
- $f'_1 f'_2$: the number of pairs of molecules with velocity \mathbf{v}'_1 and \mathbf{v}'_2 , respectively colliding in the volume d^3v at time t (incoming molecules)
- Equilibrium distribution: stationary ($\frac{\partial f}{\partial t} = 0$) solution of the Boltzmann equation (the $t \rightarrow \infty$ limit of the time dependent solutions)
- *H*-theorem: Let $H[f] = \int f \ln f d^3v$. Then
 - (i) for any solution of the Boltzmann equation $\frac{H[f]}{dt} \leq 0$, and
 - (ii) $\frac{dH[f]}{dt} = 0$ iff f is the Maxwell-Boltzmann distribution
- Sketch of the proof:
 - (i) In $\frac{dH[f]}{dt}$ the term $(f'_1 f'_2 - f_1 f_2)(\ln f_1 f_2 - \ln f'_1 f'_2)$ is never positive
 - (ii) $\frac{dH[f]}{dt}$ is zero iff $f_1 f_2 = f'_1 f'_2$ which is equivalent to $\ln f_1 + \ln f_2 = \ln f'_1 + \ln f'_2$. This is a conservation law, but the only conserved quantities are energy (proportional to \mathbf{v}^2) and momentum (proportional to \mathbf{v}). Hence $\ln f = A(\mathbf{v} - \mathbf{v}_0)^2 + \ln C$. Fixing the constants one gets the Maxwell-Boltzmann distribution
- Fine-grained Boltzmann entropy: $S_{B,f}[f] := -H[f]$

4.2 Objections

- Instrumentalist objections (Mach, Duhem)
- Why to reduce thermodynamics on *classical mechanics*?
- *Umkehrinwand* (Reversibility objection – Loschmidt, 1876)
 - Γ -space approach
 - Premise 1. Time reversal invariance of classical mechanics: if $x(t)$ is a solution, then $T(x(t))$ also is a solution
 - Premise 2. H is independent of the direction of velocity of the molecules: $H[f] = H[Tf]$
 - Conclusion. H is increasing
 - There is no a priori reason to believe that the time-reversed states do not occur in nature
 - The reversibility objection points to crucial role that the assumption of molecular chaos plays in the derivation of the *H*-theorem: for the theorem to hold the assumption of molecular chaos has to hold *at each moment of the dynamical evolution*. But whether this is the case depends on the dynamics
 - Boltzmann: "Go ahead and reverse the momenta!"
- Poincaré's recurrence theorem (1889): In a Hamiltonian system enclosed in a finite volume with a finite energy almost all states will eventually return to as close to the the initial state as one specifies
 - Poincaré's theorem: Let $(\Gamma, \mathcal{F}, \mu, \phi_t)$ 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: $B := \{x \in \Gamma \mid x \in A \ \& \ \forall t > \tau : \phi_t x \notin A\}$. Then $\mu(B) = 0$
 - 'Proof': (i) Choose a large enough t such that $\mu(B \cap B_t) = 0$. It follows that for any n and $n + k$ $\mu(B_{nt} \cap B_{(n+k)t}) = 0$. From Liouville's theorem $\mu(B_{nt}) = \mu(B_{(n+k)t}) = 0$ for any n and k . But then $\mu(B) = 0$

- *Wiederkehrinwand* (Recurrence objection – Zermelo, 1896)
 - After a Poincaré cycle the system gets back 'very' close to the initial state
 - If H is continuous (!) then it cannot decrease monotonically
 - Poincaré:
 - ”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: ”little patience” = $10^{10^{19}}$ seconds for a cubic centimetre of air (microstates are defined here within 10 \AA spatial position for each molecule and within 1-3% speed value)

4.3 Boltzmann, 1877b

- A combinatorial (non-dynamical) approach
 - coarse-graining the μ -space
 - ω_i : the i^{th} cell in the μ -space
 - $\delta\omega = \delta^3x \delta^3v$: the volume of the cells (6 dimensional cubelets) – 'large and small'
 - Three key notions:
 - * Arrangement: the specification of *which* of the n molecules is in which cell
 - * Macrostate or state distribution $Z_{\{n_i\}} = (n_1, n_2 \dots n_k)$: the specification of *how many* molecules are in which cell
 - * Microstate: $x \in \Gamma$
 - Macrostates supervene on the microstates:
 - * Microstate \rightarrow arrangement \rightarrow state distribution
 - * State distribution $\not\rightarrow$ arrangement $\not\rightarrow$ microstate: $Z_{\{n_i\}}$ remains the same if (i) we permute molecules in different cells (different arrangement), (ii) we move the molecules within the cells (same arrangement)
 - Macroregion: $\Gamma_{Z_{\{n_i\}}} := \{x \in \Gamma_E \mid Z = Z_{\{n_i\}}\}$
 - Macroregions form a partition of Γ : $\{\Gamma_{Z_{\{n_i\}}}\}$
 - Since $|\Gamma_{Z_{\{n_i\}}}| = \frac{N!}{n_1!n_2!\dots n_k!}(\delta\omega)^N$, the regions $\Gamma_{Z_{\{n_i\}}}$ vary enormously in size: a lot of molecules in the same cell $\rightarrow |\Gamma_{Z_{\{n_i\}}}|$ is small; evenly distributed arrangements $\rightarrow |\Gamma_{Z_{\{n_i\}}}|$ is big
 - Equilibrium: the distribution corresponding to the largest region (largest 'number' of different microstate)
 - With the constraints on Z : $\sum_{i=1}^l n_i = N$, $\sum_{i=1}^l n_i \epsilon_i = E$, using the method of Lagrange multipliers and the Stirling formula, Boltzmann shows that $n_i = \alpha e^{-\beta \epsilon_i}$ (which is the discrete Maxwell-Boltzmann distribution)
- Coarse-grained Boltzmann entropy: $S_{B,c}[Z_{\{n_i\}}] := k \ln |\Gamma_{Z_{\{n_i\}}}|$
 - With the Stirling formula: $S_{B,c}[Z_{\{n_i\}}] = \sum_i n_i \ln n_i + c$, or using the probability $p_i := \frac{n_i}{N}$: $S_{B,c}[Z_{\{n_i\}}] = \sum_i p_i \ln p_i + c'$
 - Entropy: (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
 - Taking the logarithm gives an extensive quantity

- The largest entropy, by definition, is assigned to the equilibrium distribution
- Remarks:
 - No mention of dynamics, collisions etc.
 - It works only for systems with no interaction (The energy of the molecules depended only on its coarse-grained microstate and not on the state of other particles)
 - Is coarse-graining subjective, objective?
 - Taking energy cells the combinatorial argument fails to reproduce the Maxwell-Boltzmann distribution
 - Cells cannot be eliminated: in the $\delta\omega \rightarrow 0$ limit, the formulae diverge
 - * Occupation number and distribution: $n_i(t) = \int_{\omega_i} f(\mathbf{r}, \mathbf{v}, t) d^3r d^3v = f(\mathbf{r}_i, \mathbf{v}_i, t) \delta\omega$ (by the mean value theorem)
 - * $S_{B,c}[Z_{\{n_i\}}] = -\sum_i \delta\omega f(\mathbf{r}_i, \mathbf{v}_i, t) \ln f(\mathbf{r}_i, \mathbf{v}_i, t) - n \ln(\delta\omega) + c \approx S_{B,f}[f] - n \ln(\delta\omega) + c$
 - The definition of equilibrium is changed
 - Evolution towards equilibrium: Does the overwhelmingly large phase space volume of Γ_{eq} compared to the set of non-equilibrium points provide a sufficient argument?

4.4 Other contributions of Boltzmann

- Ergodicity
 - Boltzmann, 1868a: first application of an ensemble instead of an individual system. Time average and phase space average: When are they equal?
 - Boltzmann, 1871: harmonic oscillator with potential $U(x, y) = ax^2 + by^2$ where $\sqrt{\frac{a}{b}}$ is irrational \rightarrow ergodic hypothesis
- Universe, entropy, arrow of time
 - Boltzmann, 1868a: Our observable world is only a small portion of a vast Universe in thermal equilibrium which came about by random fluctuation.
 - Weak antropic argument: non-equilibrium is essential for the existense of a sentient being
 - Arrow of time := the gradient of entropy \rightarrow The arrow of time (7.8)

5 Gibbs' statistical mechanics

- Gibbs, 1902: ensemble instead of individual system

"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 we know about a body can generally be described most accurately and most simply by saying that it is one taken at random from a great number (ensemble) of bodies which are completely described"

- Schrödinger: Ensembles are "mental copies of the one system under consideration", they do not interact with each other

- Probability density function: $\rho_t(x)$ ($x \in \Gamma$)
 - Liouville's theorem: $\frac{d\rho}{dt} = \frac{\partial\rho}{\partial t} + \{\rho, H\} = 0$: ρ 'flows' as an incompressible fluid²
 - ρ is defined in the Γ -space and normed to 1, f is defined in the μ -space and normed to N
 - Dynamics: $\rho_t(x) = \rho_0(\phi_{-t}x) \iff \frac{\partial\rho_t(x)}{\partial t} = \{H, \rho\}$
- Stationary density function: $\forall t, x : \rho_0(\phi_{-t}x) = \rho_0(x) \iff \frac{\partial\rho_t(x)}{\partial t} = 0$
- They characterize equilibrium since they yield constant averages: $\frac{\partial\langle f(x) \rangle}{\partial t} = \int_{\Gamma} f(x) \frac{\partial\rho_t(x)}{\partial t} dx = 0$ (if f does not depend explicitly on time)
- Special cases:

$$\rho_E(x) = \frac{1}{\omega(E)} \delta(H(x) - E) \quad (\text{microcanonical: } E, N \text{ is fixed})$$

$$\rho_T(x) = \frac{1}{Z(T)} e^{-\frac{H(x)}{T}} \quad (\text{canonical: } N \text{ is fixed, } E \text{ can vary})$$

$$\rho_{T,\alpha}(x, N) = \frac{1}{N!Z(T, \alpha)} e^{-\frac{H(x)}{T+\alpha N}} \quad (\text{grand-canonical: } E, N \text{ can vary})$$

- Postulate of equal a priori probability: all microstates compatible with the appropriate constraints have equal probability
- Fine-grained Gibbs entropy: $S_{G,f}[\rho] := -k \int_{\Gamma} \rho(x) \ln \rho(x) dx$
 - $S_{G,f}[\rho] \neq S_{B,f}[f]$: the Gibbs entropy is defined on Γ , the Boltzmann entropy is defined on the μ -space
 - For the above constraints the appropriate distribution has the maximal fine-grained Gibbs entropy
- How $S_{G,f}[\rho]$ changes with the dynamics ϕ_t ?
 - Due to Liouville's theorem the fine-grained Gibbs entropy remains constant: $\frac{\partial S_{G,f}[\rho]}{\partial t} = 0$ (See (Zeh, 2007, 52))
 - But! Gibbs' 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!)
- Coarse-graining: $\mathcal{CG} : \rho(x) \mapsto \hat{\rho}(x) = \sum_i \hat{\rho}(i) \mathbf{1}_{\omega_i}(x)$ where $\hat{\rho}(i) := \frac{1}{\delta\omega} \int_{\omega_i} \rho(x) dx$
- Coarse-grained Gibbs entropy: $S_{G,c}[\rho] := -k \int_{\Gamma} \hat{\rho}(x) \ln \hat{\rho}(x) dx$
 - $S_{G,c}[\rho]$ is not constrained by Liouville's theorem!
- Does the coarse-grained Gibbs entropy evolve towards the equilibrium state?
 - It is true that: $S_{G,c}[\rho] \geq S_{G,f}[\rho]$ (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*: a dynamical system $(\Gamma, \mathcal{F}, \mu, \phi_t)$ is called *mixing* iff for all $A, B \in \mathcal{F}$: $\lim_{t \rightarrow \infty} \mu(\phi_t A \cap B) = \mu(A)\mu(B)$
- Remarks:
 - For equilibrium systems $S_{G,f}[\rho] = S$ (thermodynamic entropy)
 - (Unlike S) $S_{G,f}[\rho]$ and $S_{G,c}[\rho]$ are properties of the ensemble

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

- $S_{G,f}[\rho]$ does *not* but $S_{G,c}[\rho]$ *does* vary with time
- Where does coarse-graining come from? Limited capacity of the observer?
- Does equilibrium make sense to creatures with unlimited capacity? → spin-echo experiment (Hahn, 1950): "A closed system in thermodynamical equilibrium may preserve an arrow of time in the form of *hidden correlations*." (Zeh, 2007, 57)
- → The many face of entropy (7.1)

6 The Ehrenfests, 1912

- A critical review of Boltzmann's and Gibbs' approach

6.1 Ergodic problem

- Three problems in Boltzmann's approach
 1. "Probability" is sometimes used as the relative volume of the region and sometimes the relative time spent in the region by the trajectory of the system
 2. Why microcanonical distribution?
 3. Why microstates evolve into Γ_{eq} ?
- Time average: $\bar{f}(x) := \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(\phi_t x) dt$ – empirically meaningful but hard to calculate
- Ensemble average: $\langle f \rangle_t := \int_{\Gamma_E} f(x) d\mu_E(x) = \int_{\Gamma_E} f(x) \rho_t(x) dx$ – manageable but what does it mean?
- Ergodic problem: When are the two averages equal?
 - Observe that $\bar{f}(x)$ depends on x and $\langle f \rangle_t$ depends on t
 - For stationary measure $\langle f \rangle_t$ does not depend on t . Take for example $\langle f \rangle_{mc}$
- Boltzmann's ergodic hypothesis: the trajectory of the system traverses all points on the energy hypersurface Γ_E , that is for any $x, y \in \Gamma_E$ there is a τ such that $y = \phi_\tau x$
 - The ergodic hypothesis solves the ergodic problem:

$$\begin{aligned} \bar{f}(y) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(\phi_{t+\tau} x) dt = \lim_{T \rightarrow \infty} \frac{1}{T} \left(\int_0^T f(\phi_t x) dt - \int_0^\tau f(\phi_t x) dt \right) = \\ &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(\phi_t x) dt = \bar{f}(x) \end{aligned}$$

- Therefore \bar{f} is constant over Γ_E , and hence $\bar{f} = \langle f \rangle_{mc}$
- → Ergodicity (7.2)

6.2 Concentration curve

- The Boltzmann equation and the H -theorem is valid not for a single system but for a *set* of systems: they are probabilistic claims
- The vast majority of systems will remain close to the solution of the Boltzmann equation and hence the H -function (*concentration curve, H-bundle*)
- Individual systems may and will deviate from the concentration curve in accord with Poincaré's recurrence theorem → The Kac ring model (7.3)

7 Topics for further discussion

7.1 The many face of entropy

- Fine-grained Boltzmann entropy: $S_{B,f}[f] := - \int f(v) \ln f(v) d^3v$
- Coarse-grained Boltzmann entropy: $S_{B,c}[Z_{\{n_i\}}] := k \ln \left| \Gamma_{Z_{\{n_i\}}} \right|$
- Fine-grained Gibbs entropy: $S_{G,f}[\rho] := -k \int_{\Gamma} \rho(x) \ln \rho(x) dx$
- Coarse-grained Gibbs entropy: $S_{G,c}[\rho] := -k \int_{\Gamma} \hat{\rho}(x) \ln \hat{\rho}(x) dx$
- Shannon entropy: $S_{Sh} := \text{Exp}(-\ln(p(X)))$ where $p(X)$ is the probability mass function of the discrete random variable X
- Kolmogorov Sinai entropy of the dynamical system $(\Gamma, \mathcal{F}, \mu, \phi)$: Let Q be a partition of Γ ; define the entropy of Q as $H(Q) := - \sum_i \mu(Q_i) \ln \mu(Q_i)$ and consider the refinement of an iterated ϕ -pullback of Q : $Q^N := \bigvee_{n=0}^N \phi^{-n}Q$. The Kolmogorov Sinai entropy is then the maximal entropy of these refinements: $S_{KS} := \sup_Q \left(\lim_{N \rightarrow \infty} \frac{1}{N} H(Q^N) \right)$ where the supremum is taken over all finite measurable partitions
- von Neumann entropy: $S_{vN} := -\text{Tr}(\rho \ln \rho)$ where ρ is a density operator

7.2 Ergodicity

- Birkhoff theorem. Let $(X, \mathcal{F}, \mu, \phi_t)$ be a dynamical system and let f be an integrable phase function. Then the time average $\bar{f}(x)$ (i) exists almost everywhere, (ii) is invariant: $\bar{f}(x_0) = \bar{f}(\phi_t x_0)$ for all t , and integrable: $\int_X \bar{f}(x_0) d\mu = \int_X f d\mu$.
- Ergodicity. A dynamical system $(X, \mathcal{F}, \mu, \phi_t)$ is *ergodic* iff for any measurable set $A \in \mathcal{F}$ such that $\mu(A) \neq 0$ and for all $x \in X$, except for a set of measure 0, it holds that $\{\phi_t(x)\} \cap A \neq \emptyset$ for some t .
- Ergodic theorem. Let $(X, \mathcal{F}, \mu, \phi_t)$ be a dynamical system and let f be an integrable phase function. Then $\bar{f} = \langle f \rangle$ iff the system is ergodic.
- Corrolaries:
 - Let f be χ_A . Then $\mu(A) = \bar{\chi}_A = \langle \chi_A \rangle = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \chi_A(\phi_t x) dt$: the measure of a region is the fraction of time the system spends in the region: a good justification of the volume measure
 - Hence, almost all trajectories come arbitrary close to any points on the energy hypersurface infinitely many times
 - Let $(X, \mathcal{F}, \mu, \phi_t)$ be ergodic. Let μ' be a ϕ_t -invariant measure that is absolutely continuous with respect to μ (that is for any measurable $A \subseteq X$ if $\mu(A) = 0$, then $\mu'(A) = 0$). Then $\mu' = \mu$. (Loosely: if the system is ergodic, there is only one time invariant measure)
- Decomposability. A dynamical system $(X, \mathcal{F}, \mu, \phi_t)$ is *decomposable* iff X can be partitioned into two (or more) invariant regions of non-zero measure, that is if there are $A, B \subseteq X$ such that $A \cap B = \emptyset$, $A \cup B = X$, $\mu(A) \neq 0 \neq \mu(B)$ and $\phi_t(A) \subseteq A$ and $\phi_t(B) \subseteq B$ for all t
 - Indecomposability \iff There is no (neglected) global constant of motion that differs in its value on sets of trajectories of non-zero measure
 - Ergodicity \iff Indecomposability
- Ergodic hierarchy: a dynamical system $(\Gamma, \mathcal{F}, \mu, \phi_t)$ is
 - *weak mixing* if for all $A, B \in \mathcal{F}$: $\lim_{t \rightarrow \infty} \frac{1}{T} \int_0^T |\mu(\phi_t A \cap B) - \mu(A)\mu(B)| dt = 0$,

- *mixing* if for all $A, B \in \mathcal{F}$: $\lim_{t \rightarrow \infty} \mu(\phi_t A \cap B) = \mu(A)\mu(B)$,
- a *K system* if there is a subalgebra $\mathcal{F}_0 \subset \mathcal{F}$ such that
 1. $\phi_n \mathcal{F}_0 \subset \phi_m \mathcal{F}_0$ for $m < n$, where ϕ_n is a discrete dynamics;
 2. the smallest σ -algebra containing $\cup_{n=1}^{\infty} \phi_{-n} \mathcal{F}_0$ is \mathcal{F} ; and
 3. $\cap_{n=1}^{\infty} \phi_n \mathcal{F}_0 = \mathcal{N}$, where \mathcal{N} is the σ -algebra containing only sets of μ -measure zero or one.
- a *Bernoulli system* if there is a finite partition $\mathcal{P} = \{A_1, A_2 \dots A_n\}$ of Γ such that the doubly infinite coarse-grained histories are isomorphic to a Bernoulli scheme for \mathcal{P} with the distribution $p_i = \mu(A_i)$ ($i = 1 \dots n$).
- Relations: Bernoulli system \implies K system \implies Mixing \implies Weak mixing, Indecomposability
 - Mixing \implies Indecomposability: Let $A = B$ invariant subsets of Γ . Then $\lim_{t \rightarrow \infty} \mu(\phi_t A \cap A) = \mu(A) = \mu(A)\mu(A) \implies \mu(A) \in \{0, 1\}$.
 - Indecomposability $\not\implies$ Mixing: 1-dimensional harmonic oscillator is decomposable but not mixing.

7.3 The Kac ring model, 1956

- N : the number of balls on the ring, revolving counterclockwise changing color at leaving active sites
- $W(t), B(t)$: the number of white, black balls at time t
- $\delta(t) := \frac{W(t) - B(t)}{N}$: normalized difference (greyness)
- A : the number of the 'active' sites
- $W_A(t), B_A(t)$: the number of white, black balls in active sites at time t
- Dynamics:

$$\begin{aligned} W(t+1) &= W(t) - W_A(t) + B_A(t) \\ B(t+1) &= B(t) - B_A(t) + W_A(t) \end{aligned}$$

- With the *Stosszahlansatz*: $\frac{W_A(t)}{W(t)} = \mu := \frac{A}{N}$ one gets for the dynamics: $\delta(t) = (1 - 2\mu)^t \delta(0)$ (**)
- The macrostate $\delta = \frac{2W - N}{N}$ fixes the number of white and black balls and is compatible with $\binom{N}{W}$ microstates, so we can define the entropy as: $S(\delta) := k \ln \binom{N}{W}$
- Using Stirling's formula: $S(\delta) \approx -\frac{1+\delta}{2} \ln \frac{1+\delta}{2} - \frac{1-\delta}{2} \ln \frac{1-\delta}{2}$ and is monotonically increasing with t (since δ is monotonically decreasing)
- Reversibility objection: in the reverse process (balls revolving clockwise and changing color at *entering* an active site) δ is monotonically decreasing again (check!) but (for example starting with all white balls) it cannot
- Recurrence objection: after $2N$ time steps the original configuration has been restored. Therefore δ cannot be monotonically decreasing
- The crucial step in the derivation was the *Stosszahlansatz*! But this does not hold for the Kac ring: after the second round trip we know how many white ball will be on active sites
- Resolution: Introduce i.i.d. random variables ϵ_i for every site with μ probability being active (representing an ensemble of rings). This yields equation (**) above *for the ensemble*. In the above objections $\langle \epsilon_i \epsilon_{i+N} \rangle = 1$ so they were not i.i.d.

7.4 Malament, Zabell, 1980

- Gibbsian strategy: physical quantities are time averages of phase functions
- Problems with time average:
 - Why do we measure time averages?
 - Moreover, why do we measure infinite time averages?
 - The change of macroscopic parameters is inconsistent with the claim that what we observe are infinite time averages
- Malament, Zabell: Retaining ergodicity but no appeal to time averages
- Assumption 1. The phase function f associated with the macroscopic parameters is such that

$$\mu_{mc} \left(\left\{ x \in \Gamma_E : \left| f(x) - \int_{\Gamma_E} f d\mu_{mc} \right| \geq \varepsilon \right\} \right) \approx 0$$

- Assumption 2. $p(A) = \mu_{mc}(A)$
- Conclusion: It is very likely that the system is in a microstate for which the value of f is close to the phase average
- Vranas, 1998: ε -ergodicity

7.5 Khinchin, 1949

- Retaining time averages but no appeal to ergodicity
- Assumption 1. Thermodynamic limit: we only have to consider systems with large number of degrees of freedom, and
- Assumption 2. a special class of phase functions: the sum functions: $f(x) = \sum_{i=1}^n f_i(x_i)$
- Khinchin theorem. For all sum functions f there are positive constants k_1 and k_2 such that

$$\mu_{mc} \left(\left\{ x \in \Gamma_E : \left| \frac{\overline{f(x)} - \langle f \rangle}{\langle f \rangle} \right| \geq k_1 n^{-\frac{1}{4}} \right\} \right) \leq k_2 n^{-\frac{1}{4}}$$

- Remarks:
 - Weaker assumptions than ergodicity
 - Measurement outcomes are still associated with infinite time averages
 - There are thermodynamic quantities which cannot be characterized by sum functions
 - "Methodological paradox": for the proof the Hamiltonian has to be a sum function but then there is no interaction leading to the equilibrium \rightarrow Khinchin: short range interactions

7.6 Landford, 1976

- Aim: to derive Boltzmann's H -theorem from the Hamiltonian dynamics in the sense that the following diagram is commutative:

$$\begin{array}{ccc} x_0 & \xrightarrow{\text{Hamilton}} & x_t \\ \downarrow & & \downarrow \\ f^{x_0} & \xrightarrow{\text{Boltzmann}} & f_t^{x_0} \stackrel{?}{=} f^{x_t} \end{array}$$

where $x \in \Gamma$ and f^x is a function on the μ -space.

- Problem: f^x has to be differentiable, but $f^{x^N} := \frac{1}{N} \sum_i^N \delta^3(\mathbf{r} - \mathbf{r}'_i) \delta^3(\mathbf{v} - \mathbf{v}'_i)$ is not.
- Solution: Take the Boltzmann-Grad limit: $N \rightarrow \infty$ such that the mean free path $\frac{Nd^2}{V}$ is constant.
- Landford's theorem (Uffink, 2007, 1031)
- Remarks:
 - Because of the Boltzmann-Grad limit used in the proof, the theorem holds only for extremely dilute gases.
 - "the theorem to be stated says only that the Boltzmann equation holds for times *no larger than one-fifth of a mean free time* and hence does not suffice to justify its physically interesting applications" (Landford)
 - There is no need to rerandomize: one need to assume the *Stosszahlansatz* only *once!*

7.7 Markov process

- Two motivations to use stochastic dynamics: coarse-graining and interventionism
- A simple stochastic model: 'dog flea model' (Ehrenfests, 1907)
- Markov process: $p(x_n, t_n | x_{n-1}, t_{n-1}; x_{n-2}, t_{n-2}; \dots x_2, t_2; x_1, t_1) = p(x_n, t_n | x_{n-1}, t_{n-1})$
 - No memory
 - Time-symmetric: equivalent definition:

$$p(x_n, t_n; \dots x_1, t_1 | x_i, t_i) = p(x_n, t_n; \dots x_{i+1}, t_{i+1} | x_i, t_i) p(x_{i-1}, t_{i-1}; \dots x_1, t_1 | x_i, t_i)$$
 - Markov chain: homogeneous Markov process: $p_{yx}^{t_2 t_1} = p_{yx}^{t_2+t, t_1+t}$
 - Stochastic matrices: T such that $T_{ij} \geq 0$ and $\sum_i T_{ij} = 1$
 - Chapman-Kolmogorov equation: $p_{zx}^{31} = \sum_y p_{zy}^{32} p_{yx}^{21}$ or equivalently $T^{31} = T^{32} \circ T^{21}$
 - Master equation: $\frac{\partial p_t(x)}{\partial t} = \sum_y (W_{xy} p^t(y) - W_{yx} p^t(x))$ like the Boltzmann equation, is a first-ordered differential equation, but, unlike the Boltzmann equation, is linear.
- Entropy increase
 - The relative entropy $H(p, q) := - \sum_x p_x \ln \frac{p_x}{q_x}$ is monotonically non-decreasing: $H(Tp, Tq) \geq H(p, q)$: p and q will resemble more and more.
 - If p^{eq} is stationary: $Tp^{eq} = p^{eq}$, then Tp will more and more resemble to p^{eq} : $H(p, p^{eq}) \leq H(Tp, Tp^{eq}) = H(Tp, p^{eq})$.
 - If p^{eq} is even uniform: $p_x^{eq} = \frac{1}{n}$, then also the absolute entropy $H(p) := - \sum_x p_x \ln p_x$ will increase, since $H(p, p^{eq}) = H(p) - \ln n$.
- Is there always a stationary distribution?
 - In the finite case, due to the Peron-Frobenius theorem, there is: every stochastic matrix has an eigenvector with exclusively non-negative components and eigenvalue 1.
 - In the infinite case, there is not always. Counter-example: Wiener process. But for continous variables with a range that has a finite measure, there is.
 - T is *irreducible*, iff it is not completely or incompletely reducible, or equivalently, if for any x, y there is a t such that $p_{xy}^t > 0$. If T is irreducible, there exists a *unique* stationary distribution. Furthermore, the following theorem holds:

- * Ergodic theorem for Markov processes: If the transition probability T_t is irreducible, the time average of p_t converges to the unique stationary solution: $\lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_0^\tau T_t p_x dt = p_x^{eq}$.
- If the system is not *asymptotically periodic*, then the monotonic behaviour of $H(p^t)$ implies that $\lim_{t \rightarrow \infty} p^t = p^{eq}$.
- Motivating Markov property by coarse-graining
 - Transition probabilities:

$$T_{ji}^t := p(x_t \in \omega_j | x_0 \in \omega_i) = p(\phi_t x_0 \in \omega_j | x_0 \in \omega_i) = p(x_0 \in \phi_{-t} \omega_j | x_0 \in \omega_i) = \frac{\int_{(\phi_{-t} \omega_j) \cap \omega_i} \rho(x) dx}{\int_{\omega_i} \rho(x) dx}$$
 - Coarse-grained dynamics: $\hat{\phi}_t(\rho(x)) := \sum_{ij} T_{ji}^t \frac{\int_{\omega_j} \rho(x) dx}{\int_{\omega_j} dx} \mathbf{1}_{\omega_j}(x)$
 - Chapman–Kolmogorov equation: $\hat{\phi}_{t'+t} = \hat{\phi}_{t'} \circ \hat{\phi}_t$
- Motivating Markov property by interventionism
 - System + environment: $\rho(x, y)$
 - Hamiltonian: $H = H^s + H^e + \lambda H^i$
 - Coarse-grained dynamics: $\hat{\phi}_t : \rho_0^s(x) \rho_0^e(y) \xrightarrow{H} \rho_t(x, y) \xrightarrow{CG} (\int \rho_t(x, y) dy) \rho_0^e(y)$ by tracing out for the system
 - Chapman–Kolmogorov equation: $\hat{\phi}_{t'+t} = \hat{\phi}_{t'} \circ \hat{\phi}_t$

7.8 The arrow of time

- Boltzmann’s approach: (i) fluctuation out of equilibrium, (ii) weak antropic argument, (iii) equating arrow of time with entropy gradient
- Modern approach: (i) low entropy initial state
- In what sense the increase of entropy is accounting for the time arrow?
 - up-down, left-right, earlier-later
 - Reichenbach: footprint in the sand, Earman: cat food cans in the supermarket
 - Lewis: counterfactual theory of causation
 - substance identification
- Zeh, 2007

7.9 Gibbs paradox

7.10 Maxwell’s demon

7.11 Phase transition

7.12 *Stosszahlansatz* and Projection Postulate

7.13 The problems of rerandomization

7.14 Quantum statistical physics

Selected readings

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

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

Jos Uffink: "Bluff your way with the Second Law of Thermodynamics," *Studies in the History and Philosophy of Modern Physics*, **32**, 305-394.

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

Roman 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).

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

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