3

The Thermodynamical Arrow of Time

The thermodynamical arrow of time is characterized by the increase of entropy according to the Second Law. This law was first postulated by Rudolf Clausius in 1865 as a consequence of Carnot's theorem of 1824 when combined with the just established equivalence of heat with other forms of energy (the First Law of thermodynamics). It can be written in a general form by means of a sum of external and internal changes of entropy as

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \left\{\frac{\mathrm{d}S}{\mathrm{d}t}\right\}_{\mathrm{ext}} + \left\{\frac{\mathrm{d}S}{\mathrm{d}t}\right\}_{\mathrm{int}} ,$$

where

$$\mathrm{d}S_{\mathrm{ext}} = \frac{\mathrm{d}Q}{T} \quad \text{and} \quad \left\{\frac{\mathrm{d}S}{\mathrm{d}t}\right\}_{\mathrm{int}} \geq 0 . \tag{3.1}$$

Here, S is phenomenologically defined as the entropy of a bounded system – thereby exploiting reversible processes with $(\mathrm{d}S/\mathrm{d}t)_{\mathrm{int}} = 0$, while $\mathrm{d}Q$ is the reversible (infinitely slow) inward heat flux through the system's complete boundary during a time interval dt. (See also the local form (3.39) of the Second Law on p. 60.)

Conventionally, the heat flux is not written as a derivative $\mathrm{d}Q/\mathrm{d}t$, since its integral $Q(t)$ would not represent a 'function of state' – although it does, of course, define the time-integrated net flux in the actual process. The first term of $\mathrm{d}S/\mathrm{d}t$ in (3.1) vanishes by definition for 'thermodynamically closed' systems. Since the whole Universe is defined as an absolutely closed system (even if infinite), its total entropy, or the mean entropy of co-expanding volume elements, should according to this law evolve towards its maximum – the so-called *Wärmetod* (heat death) of the world. The phenomenological thermodynamical concepts used in (3.1), in particular the temperature, apply only in situations of partial (local) equilibrium.

Statistical physics is now believed to provide an explanation and potential generalization of phenomenological thermodynamics – including its Second Law. While in principle *all* physical concepts are phenomenological, this term is used here to emphasize the presumed existence of conceivably complete microscopic concepts that await the application of statistical methods.

While statistical considerations are indeed essential for the understanding of thermodynamical concepts, statistics as a method of counting has nothing a priori to do with dynamics. Therefore, it cannot by itself explain dynamically 'irreversible' processes – characterized by $\{\mathrm{d}S/\mathrm{d}t\}_{\mathrm{int}} > 0$. This requires *additional* assumptions, which often remain unnoticed, since they appear 'natural' to our prejudiced way of thinking in terms of 'causes' (exclusively in the past). These hidden assumptions have therefore to be carefully investigated in order to reveal the true origin of the thermodynamical arrow.

An attempt to explain this fundamental asymmetry on the basis of the 'historical nature' of the world, that is, by using the idea that the past is 'fixed' (and therefore neither requires nor allows statistical retrodiction) would clearly represent a circular argument when starting from nothing but time-symmetrically deterministic laws. This idea must itself be rooted in the time asymmetry of the physical world. The existence of reliable knowledge or information only about the past corresponds to a time-asymmetric *physical* relation between documents and their sources, analogous to the asymmetric 'causal' relation between retarded electromagnetic fields and charged currents as their advanced sources (Sect. 2.1). For example, light contains information about objects in the more or less recent past. Similarly, *all* documents represent an asymmetry in the physical world, and do not simply reflect the way boundary conditions (such as initial or final conditions) are posed.

In a statistical description, 'irreversible' processes are of the form

$$\text{improbable state} \underset{t}{\longrightarrow} \text{probable state}\,,$$

where the probability ratio is usually a huge number. These probabilities are defined by the size or measure of certain *sets* of elementary states (called 'representative ensembles' by Tolman 1938), which contain the *real* state of the considered system (a point in its microscopic configuration space) as a member. This measure of probability changes as the state moves along its trajectory through different such sets. If the representative ensembles are *operationally* defined, for example by means of macroscopic preparation procedures, they are often themselves called macroscopic or thermodynamical 'states'. This terminology has its origin in a description that is unaware of the microscopic states (or rejects the concept of such a microscopic reality). The dynamical justification of thermodynamical states is a major objective of a microscopic foundation of thermodynamics: why are certain sets of microscopic states 'representative' in forming macroscopic states?

Irreversible processes of the above kind would statistically be more abundant than those of the kind

$$\text{improbable state} \underset{t}{\longrightarrow} \text{improbable state}\,.$$

Their overwhelming occurrence in Nature can therefore be understood under the presumption of improbable *initial* states. In an operational approach, such an assumption would simply be taken for granted: a consequence of operations to be performed in time. In a cosmological context it requires a cosmic initial condition. This has occasionally been called the *Kaltgeburt* (cold birth) of the Universe, although a low temperature (kT much smaller than energies of mechanical degrees of freedom) need not be its essential aspect – see Sect. 5.3. However, this initial assumption appears quite unreasonable precisely for statistical reasons, since (1) there are just as many processes of the type

$$\text{probable state} \underset{t}{\longrightarrow} \text{improbable state}\,,$$

and (2) far more of the kind

$$\text{probable state} \underset{t}{\longrightarrow} \text{probable state}\,.$$

The latter describe equilibrium. Hence, for statistical reasons we should expect the world to *be* in the situation of a heat death, while the required improbable initial condition needs an explanation that does *not presume* causality.

The first of these two arguments is the 'reversibility objection' (*Umkehreinwand*), formulated by Boltzmann's friend and teacher Johann Joseph Loschmidt. It is based on the fact that each trajectory has precisely one time-reversed counterpart.[1] If, for example, $z(t) \equiv \{q_i(t), p_i(t)\}_{i=1,\ldots,3N}$ describes a trajectory in $6N$-dimensional phase space (*Γ-space*) according to the Hamiltonian equations, then the time-reversed trajectory, $z_{\mathrm{T}}(-t) \equiv \{q_i(-t), -p_i(-t)\}$, is also a solution of the equations of motion. If the entropy S of a state z can be defined as a function of this state, $S = F(z)$, with $F(z) = F(z_{\mathrm{T}})$, then Loschmidt's objection means that for every solution with $\mathrm{d}S/\mathrm{d}t > 0$ there is precisely one corresponding solution with $\mathrm{d}S/\mathrm{d}t < 0$. In statistical theories, $F(z)$ is defined as a monotonic function (conveniently the logarithm) of the size or measure of the mentioned set of states to which z belongs. The property $F(z) = F(z_{\mathrm{T}})$ is then a consequence of the symmetry character of the transformation $z \to z_{\mathrm{T}}$, while the stronger objection (2) above means that there are far more solutions with $\mathrm{d}S/\mathrm{d}t \approx 0$, that is, $S(t) \approx S_{\max}$ – simply because this condition characterizes almost all of configuration space.

In order to justify the thermodynamical arrow of time statistically, one therefore has to either derive the improbable initial conditions from an independent (time-asymmetric) cosmological assumption, or simply postulate them in some form. The Second Law is by no means *incompatible* with deterministic or T-symmetric dynamical laws; it is just extremely improbable and

[1] Often, T (or CPT) symmetry of the dynamics is assumed for this argument. This has misleadingly given rise to the by no means justified expectation that the difficulties in deriving the Second Law may be overcome by dropping this symmetry. However, as already pointed out in the Introduction, the crucial point in Loschmidt's argument is the time reversal symmetry of *determinism* itself (not of its precise form), which is often reflected by the possibility of compensating time reversal by another symmetry operation (see also Sect. 3.4).

in conflict with unbiased statistical reasoning. The widespread 'double standard' of readily accepting improbable initial conditions while rejecting similar final ones has been duly criticized by Price (1996).

Another argument against the statistical interpretation of irreversibility, the *recurrence objection* (or *Wiederkehreinwand*), was raised much later by Ernst Friedrich Zermelo, a collaborator of Max Planck at a time when the latter still opposed atomism, and instead supported the 'energeticists', who attempted to understand energy and entropy as fundamental 'substances'. This argument is based on a mathematical theorem due to Henri Poincaré, which states that every bounded mechanical system will return as close as one wishes to its initial state within a sufficiently large time. The entropy of a closed system would therefore have to return to its former value, provided only the function $F(z)$ is continuous. This is a special case of the *quasi-ergodic theorem* which asserts that every system will come arbitrarily close to any point on the hypersurface of fixed energy (and possibly with other fixed analytical constants of the motion) within finite time.

While all these theorems are mathematically correct, the recurrence objection fails to apply to reality for quantitative reasons. The age of our Universe is much smaller than the Poincaré recurrence times even for a gas consisting of no more than a few tens of particles. Their recurrence to the vicinity of their initial states (or their coming close to any other similarly specific state) can therefore be excluded in practice. Nonetheless, some 'foundations' of irreversible thermodynamics in the literature rely on formal idealizations that would lead to *strictly infinite* Poincaré recurrence times (for example the 'thermodynamical limit' of infinite particle number). Such assumptions are not required in our Universe of finite age, and they would *not* invalidate the reversibility objection (or the equilibrium expectation, mentioned above). However, all foundations of irreversible behavior have to presume some very improbable initial conditions.

The theory of thermodynamically irreversible processes must therefore address two main problems:

1. The investigation of realistic *mechanisms* which describe the dynamical evolution away from certain (presumed) improbable initial states. This is usually achieved in the form of 'master equations', which mimic a *law-like* T-asymmetry – analogous to Ritz's retarded action-at-a-distance in electrodynamics. In contrast to electrodynamics, they describe the dynamics of ensembles, equivalent to an effective stochastic dynamics for the individual states (applicable in the 'forward' direction of time). These mechanisms should be able to justify the *representative ensembles* (macroscopic states) and even describe the emergence of *order* (Sect. 3.4).
2. The precise nature of the required improbable initial states. This leads again to the quest for an appropriate *cosmic* initial condition, similar to the global condition $A^{\mu}_{\text{in}} = 0$ in the early Universe that would be able to explain the radiation arrow (see Sects. 2.2 and 5.3).

3.1 The Derivation of Classical Master Equations

Statistical physics is concerned with systems consisting of a large number of microscopic constituents which are known to obey quantum mechanics. However, quantum theory is still haunted by interpretational problems, in particular regarding the nature of probabilistic 'quantum events'. These are usually understood as representing a *fundamental* irreversible part of dynamics, that might even be the true source of thermodynamical irreversibility. In contrast, classical mechanics is deterministic and well defined. Therefore, *classical* statistical mechanics will be formulated and discussed in this chapter for conceptual consistency and later comparison with quantum statistical mechanics – even though it is based on an incorrect microscopic theory. Most thermodynamic properties of a gas, for example, can in fact be modelled by a system of interacting classical mass points – see (4.21). While the present section follows historical routes, a more general and systematic formalism, that can later also be used in quantum theory, will be presented in Sect. 3.2.

3.1.1 μ-Space Dynamics and Boltzmann's H-Theorem

The complete *dynamical state* of a mechanical system of N classical particles (distinguishable mass points) can either be represented by one point in its $6N$-dimensional phase space ('Γ-space'), or by N numbered points in six-dimensional 'μ-space' (the single-particle phase space). These N points form a *discrete distribution* in μ-space. If the particles are *not* distinguished from one another, this is exactly equivalent to an ensemble of $N!$ points in Γ-space that results from all particle permutations. Because of the large number of particles forming macroscopic systems (of order 10^{23}), Boltzmann (1866, 1896) used continuous (smoothed) distributions (or *phase space densities*) $\rho_\mu(\boldsymbol{p}, \boldsymbol{q})$ to describe them. This plausible approximation will turn out to have important consequences.

Two types of argument are in general used to justify it:

1. The formal *thermodynamical limit* $N \to \infty$. This represents an idealization that would lead to infinite Poincaré recurrence times. Mathematical proofs may then appear rigorous, while in fact they are approximations – valid only for the early (far from equilibrium) stage of our Universe. Though often convenient, this procedure may conceal physically important aspects, in particular when interchanging the thermodynamical limit with the limit $t \to \infty$ (physically a quantitative question).
2. Slightly 'uncertain' positions and momenta, defining small volume elements in Γ-space, $\Delta V_\Gamma = (\Delta V_\mu)^N$, instead of points. They describe infinite *ensembles* of states, and they may again lead to smooth distributions, since $N!$ such volume elements easily overlap even for a dilute gas, as $N!\Delta V_\Gamma \approx (N\Delta V_\mu)^N$ according to Stirling's approximation. Although uncertainties slightly larger than distances between the particles are sufficient for the smoothing, they will turn out to have drastic dynamical

consequences for many interacting particles. However, these uncertainties *cannot* be based on the quantum mechanical uncertainty relations with their corresponding phase space cells of size h^{3N}, since equivalent problems reappear in quantum theory if phase space points are consistently replaced by wave functions (see Sect. 4.1.1).

The time dependence of an individual point $\{p_i(t), q_i(t)\}$ in Γ-space (with $i = 1, \ldots, 3N$), described by Hamilton's equations, is equivalent to the simultaneous time dependence of all N points in μ-space. Therefore, the time dependence of an ensemble in Γ-space (represented by a distribution ρ_Γ) determines that of the corresponding density ρ_μ. In contrast to the dynamics in Γ-space (Sect. 3.1.2), however, this dynamics is not 'autonomous': the time derivative of a non-singular density ρ_μ is *not* determined by ρ_μ. The reason is that ρ_Γ cannot be recovered from ρ_μ in order to determine the latter's time derivative from that of the former. The mapping of Γ-space distributions on μ-space distributions cannot be uniquely inverted, as it destroys information about correlations between the particles (see also Fig. 3.1 and the subsequent discussion). The smooth μ-space distribution may, for example, characterize a 'macroscopic state' in the sense mentioned in the introduction to this chapter. Therefore, the envisioned chain of computation

$$\rho_\mu \longrightarrow \rho_\Gamma \xrightarrow{H} \frac{\mathrm{d}\rho_\Gamma}{\mathrm{d}t} \longrightarrow \frac{\partial \rho_\mu}{\partial t} , \tag{3.2}$$

which would be required to derive an autonomous dynamics for ρ_μ, is broken at its first link. Boltzmann's attempt to bridge this gap by statistical arguments will turn out to be the source of the time direction asymmetry in his statistical mechanics, and similarly in other formulations of irreversible processes. His procedure specifies a direction in time in a phenomenologically justified way, although it was originally meant to represent a general approximation rather than a modification of the Hamiltonian dynamics. One must then ask under what circumstances it may be valid.

Boltzmann *postulated* a stochastic dynamical law of the form

$$\frac{\partial \rho_\mu}{\partial t} = \left\{ \frac{\partial \rho_\mu}{\partial t} \right\}_{\text{free+ext}} + \left\{ \frac{\partial \rho_\mu}{\partial t} \right\}_{\text{collision}} . \tag{3.3}$$

Its first term is defined to describe particle motion under external forces only. It can be written as a continuity equation in 6-dimensional μ-space:

$$\begin{aligned} \left\{ \frac{\partial \rho_\mu}{\partial t} \right\}_{\text{free+ext}} &= -\mathrm{div}_\mu j_\mu := -\nabla_q \cdot (\dot{\boldsymbol{q}} \rho_\mu) - \nabla_p \cdot (\dot{\boldsymbol{p}} \rho_\mu) \\ &= -\nabla_q \cdot \left(\frac{\boldsymbol{p}}{m} \rho_\mu \right) - \nabla_p \cdot (\boldsymbol{F}_{\text{ext}} \rho_\mu) , \end{aligned} \tag{3.4}$$

where j_μ is the current density in μ-space. In the absence of particle interactions this equation would describe the dynamics of the 'phase space fluid' exactly. It represents the *local* conservation of probability in μ-space according to the deterministic Hamiltonian equations, which hold separately for each particle in this case. Each point in μ-space (each single-particle state) moves continuously on a trajectory that is governed by the external forces $\boldsymbol{F}_{\text{ext}}$, thereby retaining its individual probability which was determined by the initial condition for ρ_μ.

For the second (non-trivial) term, Boltzmann proposed his *Stoßzahlansatz* (collision equation), which will be formulated here for simplicity under the following assumptions:

(1)	$\boldsymbol{F}_{\text{ext}} = 0$	'no external forces',
(2)	$\rho_\mu(\boldsymbol{p}, \boldsymbol{q}, t) = \rho_\mu(\boldsymbol{p}, t)$	'homogeneous distribution'.

The second condition is dynamically consistent for translation-invariant forces. From these assumptions one obtains $\{\partial\rho_\mu/\partial t\}_{\text{free+ext}} = 0$. The *Stoßzahlansatz* is then written in the plausible form

$$\frac{\partial \rho_\mu}{\partial t} = \left\{\frac{\partial \rho_\mu}{\partial t}\right\}_{\text{collision}} = \text{gains} - \text{losses}\,, \tag{3.5}$$

that is, as a *balance equation*. Its two terms on the RHS can be explicitly written in terms of transition rates $w(\boldsymbol{p}_1\boldsymbol{p}_2; \boldsymbol{p}'_1\boldsymbol{p}'_2)$ for particle pairs scattered from $\boldsymbol{p}'_1\boldsymbol{p}'_2$ to $\boldsymbol{p}_1\boldsymbol{p}_2$. They are usually (in a low density approximation) determined by the two-particle scattering cross-section, and they have to satisfy certain conservation laws. Because of this description in terms of rates for discontinuous changes of momenta, the collisions cannot be described by a *local* conservation of probability in μ-space, as in (3.4).

This *Stoßzahlansatz* (3.5) reads explicitly

$$\frac{\partial \rho_\mu(\boldsymbol{p}_1, t)}{\partial t} = \int \Big[w(\boldsymbol{p}_1\boldsymbol{p}_2; \boldsymbol{p}'_1\boldsymbol{p}'_2)\rho_\mu(\boldsymbol{p}'_1, t)\rho_\mu(\boldsymbol{p}'_2, t) - w(\boldsymbol{p}'_1\boldsymbol{p}'_2; \boldsymbol{p}_1\boldsymbol{p}_2)\rho_\mu(\boldsymbol{p}_1, t)\rho_\mu(\boldsymbol{p}_2, t)\Big] \mathrm{d}^3p_2\mathrm{d}^3p'_1\mathrm{d}^3p'_2\,. \tag{3.6}$$

It forms the prototype of a *master equation* as an irreversible balance equation based on probabilistic transition rates. Because of their time asymmetry, these master equations cannot be generally valid approximations. They may hold for special solutions, which thus characterize an arrow of time. These solutions cannot even be particularly frequent among all other solutions.

For further simplification, invariance of the transition rates under *collision inversion*,

$$w(\boldsymbol{p}_1\boldsymbol{p}_2; \boldsymbol{p}'_1\boldsymbol{p}'_2) = w(\boldsymbol{p}'_1\boldsymbol{p}'_2; \boldsymbol{p}_1\boldsymbol{p}_2)\,, \tag{3.7}$$

will be assumed. It may be derived from invariance under space reflection *and* time reversal, although these two symmetries do not necessarily have to be separately valid. The *Stoßzahlansatz* then assumes the form

$$\frac{\partial \rho_\mu(\boldsymbol{p}_1,t)}{\partial t} = \int w(\boldsymbol{p}_1\boldsymbol{p}_2;\boldsymbol{p}_1'\boldsymbol{p}_2')\Big[\rho_\mu(\boldsymbol{p}_1',t)\rho_\mu(\boldsymbol{p}_2',t) - \rho_\mu(\boldsymbol{p}_1,t)\rho_\mu(\boldsymbol{p}_2,t)\Big]\mathrm{d}^3p_2\mathrm{d}^3p_1'\mathrm{d}^3p_2' \, . \tag{3.8}$$

In order to demonstrate the irreversibility described by the *Stoßzahlansatz*, it is useful to consider *Boltzmann's H-functional*

$$H[\rho_\mu] := \int \rho_\mu(\boldsymbol{p},\boldsymbol{q},t)\ln\rho_\mu(\boldsymbol{p},\boldsymbol{q},t)\mathrm{d}^3p\,\mathrm{d}^3q = N\overline{\ln\rho_\mu} \, , \tag{3.9}$$

proportional to the mean logarithm of probability. The mean $\bar{f}$ of a function $f(\boldsymbol{p},\boldsymbol{q})$ is defined here as $\bar{f} := \int f(\boldsymbol{p},\boldsymbol{q})\rho_\mu(\boldsymbol{p},\boldsymbol{q})\mathrm{d}^3p\,\mathrm{d}^3q/N$, in accordance with the normalization $\int \rho_\mu(\boldsymbol{p},\boldsymbol{q})\mathrm{d}^3p\,\mathrm{d}^3q = N$. Because of this fixed normalization, the H-functional is large for narrow distributions, but small for wide ones. An ensemble of discrete points (or δ-distributions), for example, would lead to $H[\rho_\mu] = \infty$, while a constant distribution on a region of volume V_μ, $\rho_\mu = N/V_\mu$, gives $H[\rho_\mu] = N(\ln N - \ln V_\mu)$. Note that H is defined only up to an additive constant that depends on the choice of a unit volume element of phase space in (3.10).

One may now derive Boltzmann's *H-theorem*,

$$\frac{\mathrm{d}H[\rho_\mu]}{\mathrm{d}t} \leq 0 \, , \tag{3.10}$$

by differentiating $H[\rho_\mu]$ with respect to time, while using the collision equation in the form (3.8):

$$\begin{aligned}\frac{\mathrm{d}H[\rho_\mu]}{\mathrm{d}t} &= V\int \frac{\partial\rho_\mu(\boldsymbol{p}_1,t)}{\partial t}\big[\ln\rho_\mu(\boldsymbol{p}_1,t)+1\big]\mathrm{d}^3p_1 \\ &= V\int w(\boldsymbol{p}_1\boldsymbol{p}_2;\boldsymbol{p}_1'\boldsymbol{p}_2')\Big[\rho_\mu(\boldsymbol{p}_1',t)\rho_\mu(\boldsymbol{p}_2',t) - \rho_\mu(\boldsymbol{p}_1,t)\rho_\mu(\boldsymbol{p}_2,t)\Big] \\ &\qquad \times \big[\ln\rho_\mu(\boldsymbol{p}_1,t)+1\big]\mathrm{d}^3p_1\mathrm{d}^3p_2\mathrm{d}^3p_1'\mathrm{d}^3p_2' \, . \end{aligned} \tag{3.11}$$

The last expression may be conveniently reformulated by using the symmetries under collision inversion given by (3.7), and under particle permutation, $w(\boldsymbol{p}_1\boldsymbol{p}_2;\boldsymbol{p}_1'\boldsymbol{p}_2') = w(\boldsymbol{p}_2\boldsymbol{p}_1;\boldsymbol{p}_2'\boldsymbol{p}_1')$. (Otherwise this combined symmetry would be required to hold for short *chains* of collisions, at least.) Rewriting the integral in (3.11) as a symmetric sum of the four equivalent permutations of the integration variables, one obtains

$$\begin{aligned}\frac{\mathrm{d}H[\rho_\mu]}{\mathrm{d}t} &= \frac{V}{4}\int w(\boldsymbol{p}_1\boldsymbol{p}_2;\boldsymbol{p}_1'\boldsymbol{p}_2')\Big[\rho_\mu(\boldsymbol{p}_1',t)\rho_\mu(\boldsymbol{p}_2',t) - \rho_\mu(\boldsymbol{p}_1,t)\rho_\mu(\boldsymbol{p}_2,t)\Big] \\ &\times\Big\{\ln\big[\rho_\mu(\boldsymbol{p}_1,t)\rho_\mu(\boldsymbol{p}_2,t)\big] - \ln\big[\rho_\mu(\boldsymbol{p}_1',t)\rho_\mu(\boldsymbol{p}_2',t)\big]\Big\}\mathrm{d}^3p_1\mathrm{d}^3p_2\mathrm{d}^3p_1'\mathrm{d}^3p_2' \leq 0 \, . \end{aligned} \tag{3.12}$$

This integrand is manifestly non-positive, since the logarithm is a monotonically increasing function of its argument. This completes the proof of (3.10), which would apply to *any* monotonic function, not just the logarithm.

In order to recognize the relation between the H-functional and entropy, one may consider the *Maxwell distribution* ρ_{M}, given by

$$\rho_{\mathrm{M}}(\boldsymbol{p}) := \frac{N}{V} \frac{\exp(-p^2/2mkT)}{\sqrt{(2\pi mkT)^3}} \ . \tag{3.13}$$

Its H-functional $H[\rho_{\mathrm{M}}]$ has two important properties:

1. It represents a *minimum for given energy*, $E = \int \rho_\mu(\boldsymbol{p})[p^2/2m]\mathrm{d}^3p \approx \sum_i \boldsymbol{p}_i^2/2m$. A proof will be given in a somewhat more general form in Sect. 3.1.2. (Statistical reasoning unconstrained by a given energy value would predict infinite energy, since the phase space volume grows non-relativistically as its $(3N/2)$ th power.) ρ_{M} must therefore represent an equilibrium distribution (with maximum entropy) under the *Stoßzahlansatz* if the transition probabilities are assumed to conserve energy.
2. One obtains explicitly

$$\begin{aligned} H[\rho_{\mathrm{M}}] &= V \int \rho_{\mathrm{M}}(\boldsymbol{p}) \ln \rho_{\mathrm{M}}(\boldsymbol{p}) \mathrm{d}^3p \\ &= -N \left(\ln \frac{V}{N} + \frac{3}{2} \ln T + \text{constant} \right) \ . \end{aligned} \tag{3.14}$$

This expression may be compared with the entropy of a mole of a monatomic ideal gas according to phenomenological thermodynamics:

$$S_{\text{ideal}}(V,T) = R \left(\ln V + \frac{3}{2} \ln T \right) + \text{constant}' \ , \tag{3.15}$$

with another constant that may depend on the particle number N according to its derivation. The second constant may then be chosen such that

$$S_{\text{ideal}} = -kH[\rho_{\mathrm{M}}] =: S_\mu[\rho_{\mathrm{M}}] \ , \tag{3.16}$$

where $k = R/N$.

The entropy of an ideal gas can thus be identified with the measure of the width of the molecular distribution in μ-space. The *Stoßzahlansatz* successfully describes the evolution of this distribution towards a Maxwell distribution with its parameter T that determines the conserved total energy. This *Lagrange parameter* – see (3.19) – is thereby recognized as the *temperature*.

This important success seems to be the origin of the 'myth' of the statistical foundation of the thermodynamical arrow of time. However, statistical arguments applied to a gas can neither explain why the *Stoßzahlansatz* is a good approximation in one and only one direction of time, nor tell us whether S_μ is *always* an appropriate definition of entropy. It will indeed turn out to be insufficient when correlations between particles become essential, as is the case, for example, for real gases or solid bodies. Taking them into account requires more general concepts, which were first proposed by Gibbs. His approach will also allow us to formulate the exact ensemble dynamics in Γ-space, although it *cannot* yet explain the origin of the thermodynamical arrow of time (that is, of the low-entropy initial conditions).

3.1.2 Γ-Space Dynamics and Gibbs' Entropy

In the preceding section, Boltzmann's *smooth* phase space density ρ_μ was justified by means of small uncertainties in particle positions and momenta. It describes an infinite number (a continuum) of *possible* single-particle states, for example each particle represented by a small volume element ΔV_μ. An objective ('real') state would instead be described by a point (or a δ-distribution) in Γ-space, or by a sum over N δ-functions in μ-space. This would then lead to an infinite value of Boltzmann's H-functional, or negative infinite entropy.

However, the finite value of $S_\mu[\rho_\mu]$, derived from the *smooth* μ-space distribution, is *not* just a measure of this arbitrary smoothing procedure (for example representing the size of the volume elements ΔV_μ). If N points are replaced by small but overlapping volume elements, this leads to a smooth distribution ρ_μ whose width reflects that of the discrete (real) distribution of particles. Therefore, S_μ characterizes the real physical state. The formal 'renormalization of entropy', which is part of this smoothing procedure, adds an infinite positive contribution to the infinite negative entropy corresponding to a point in such a way that the finite result $S_\mu[\rho_\mu]$ is *physically* meaningful. The 'representative ensemble' obtained in this way defines a finite measure of probability (in the sense of the introduction to this chapter) for the $N!$ points in Γ-space. It depends only slightly on the precise smoothing conditions, provided the discrete μ-space distribution is already smooth in the mean.

The ensemble concept introduced by Josiah Willard Gibbs (1902) differs from Boltzmann's at the very outset. He considered probability densities $\rho_\Gamma(p,q)$ with $\int \rho_\Gamma(p,q)\mathrm{d}p\,\mathrm{d}q = 1$ – from now on writing $p := p_1,\dots,p_{3N}$, $q := q_1,\dots,q_{3N}$ and $\mathrm{d}p\,\mathrm{d}q := \mathrm{d}^{3N}p\mathrm{d}^{3N}q$ for short, which are meant to describe *incomplete information* ('ignorance') about microscopic degrees of freedom. For example, a probability density may characterize a macroscopic (incomplete) preparation procedure. Boltzmann's H-functional is then replaced by Gibbs' formally analogous *extension in phase* η:

$$\eta[\rho_\Gamma] := \overline{\ln \rho_\Gamma} = \int \rho_\Gamma(p,q) \ln \rho_\Gamma(p,q)\mathrm{d}p\,\mathrm{d}q \; . \tag{3.17}$$

It leads generically to a finite *ensemble entropy* $S_\Gamma := -k\eta[\rho_\Gamma]$. For a probability density that is constant on a phase space volume element of size ΔV_Γ (while vanishing elsewhere), one has $\eta[\rho_\Gamma] = -\ln \Delta V_\Gamma$. The entropy $S_\Gamma = k \ln \Delta V_\Gamma$

is a logarithmic measure of the size of this volume element: it does not characterize a real state, as Boltzmann's entropy was supposed to do.

For a smooth distribution of statistically independent particles, $\rho_\Gamma = \prod_{i=1}^{N} \left[\rho_\mu(\boldsymbol{p}_i, \boldsymbol{q}_i)/N\right]$, one nevertheless obtains

$$\eta[\rho_\Gamma] = \sum_{i=1}^{N} \int \left[\rho_\mu(\boldsymbol{p}_i, \boldsymbol{q}_i)/N\right] \ln \left[\rho_\mu(\boldsymbol{p}_i, \boldsymbol{q}_i)/N\right] \mathrm{d}^3 p_i \mathrm{d}^3 q_i$$
$$= \int \rho_\mu(\boldsymbol{p}, \boldsymbol{q}) \left[\ln \rho_\mu(\boldsymbol{p}, \boldsymbol{q}) - \ln N\right] \mathrm{d}^3 p \, \mathrm{d}^3 q = H[\rho_\mu] - N \ln N \ . \quad (3.18)$$

In this important special case one thus recovers Boltzmann's statistical entropy S_μ (with all its advantages) – except for the term $kN \ln N \approx k \ln N!$ that has to be interpreted as the *mixing entropy* of the gas with itself. It is absent in Boltzmann's approach, since his μ-space distribution does not distinguish between particle permutations even though they define different states. While merely an additive constant in systems with fixed particle number, this self-mixing entropy leads to observable consequences *at variance with experimental results* in situations where the particle number may vary dynamically. Large particle numbers would then acquire far too large statistical weights. In particular, the specific volume V/N in (3.14) would then be replaced by the total volume V. This does even appear consistent (though empirically wrong), since particles forming an ideal gas are independent of one another, so each one is constrained only to the *total* volume V.

Since empirically not required, this self-mixing entropy was generally overlooked in Boltzmann's approach, although it had already been known as a problem to Maxwell. It can be resolved only by applying Gibbs' ensemble concept to quantum states defined in the occupation number representation for field modes (field quantization).[2] Only after borrowing this result from quantum field theory may one identify Boltzmann's entropy with an ensemble entropy (representing incomplete knowledge) for non-interacting 'particles'.

[2] The popular argument that this self-mixing entropy has to be dropped simply because of the *indistinguishability* of particles is wrong, since conceptually different (even though operationally indistinguishable) states would have to be counted separately for statistical purposes. Classical states differing by a permutation of particles would dynamically retain their individuality. The use of μ-space distributions, such as in Boltzmann's statistical mechanics, is also inconsistent from a classical point of view, unless these probability densities were multiplied by the weight factors $N!$ again. The concepts of indistinguishability and identity are different in principle (see also Saunders 2005 and references therein for a discussion). The identity of states with interchanged 'particles' can be understood in terms of quantum *fields* – see also (4.21), since the permutation of two identical wave packets at different places would represent an identity operation (Zeh 2003). Even the difference $N \ln N - \ln N! \approx N - \ln N$, usually neglected in these arguments, can be understood: it counts states with different particle numbers which must contribute to open systems that permit particle exchange, described by a *grand* canonical distribution with given chemical potential (see p. 71).

Furthermore, S_Γ is maximized under the constraint of fixed *mean* energy, $\bar{E} = \int H(p,q)\rho_\Gamma(p,q)\mathrm{d}p\,\mathrm{d}q$, by the *canonical* (or Gibbs') distribution $\rho_{\mathrm{can}} := Z^{-1}\exp\big[-H(p,q)/kT\big]$. The latter can be derived from a variational procedure with the additional constraint of fixed normalization of probability, $\int \rho_\Gamma(p,q)\mathrm{d}p\,\mathrm{d}q = 1$, that is, from

$$\delta\left\{\eta[\rho_\Gamma] + \alpha\int \rho_\Gamma(p,q)\mathrm{d}p\,\mathrm{d}q + \beta\int H(p,q)\rho_\Gamma(p,q)\mathrm{d}p\,\mathrm{d}q\right\} \tag{3.19}$$

$$= \int \Big[\ln\rho_\Gamma(p,q) + (\alpha+1) + \beta H(p,q)\Big]\delta\rho_\Gamma(p,q)\mathrm{d}p\,\mathrm{d}q = 0\ ,$$

with Lagrange parameters α and β for fixed normalization and energy. The solution is

$$\rho_{\mathrm{can}} = \exp\Big\{-\big[\beta H(p,q) - \alpha - 1\big]\Big\} =: Z^{-1}\exp\big[-\beta H(p,q)\big]\ , \tag{3.20}$$

and one recognizes $\beta = 1/kT$ and the *partition function* (sum over states) $Z := \int \mathrm{e}^{-\beta H(p,q)}\mathrm{d}p\,\mathrm{d}q = \mathrm{e}^{-\alpha-1}$. By using the *Ansatz* $\rho = \mathrm{e}^{\chi+\Delta\chi}$ with $\mathrm{e}^{\chi} := \rho_{\mathrm{can}}$, an arbitrary (not necessarily small) variation $\Delta\chi(p,q)$, the above constraints, and the general inequality $\Delta\chi\mathrm{e}^{\Delta\chi} \geq \Delta\chi$, one may even show that the canonical distribution represents an *absolute* maximum of this entropy. In statistical thermodynamics (and in contrast to phenomenological thermodynamics), entropy is thus a more fundamental concept than temperature, which applies only to special (canonical or equivalent) probability distributions, while a formal entropy is defined for *all* ensembles.

One can similarly show that S_Γ is maximized by the microcanonical ensemble $\rho_{\mathrm{micro}} \equiv \delta\big(E - H(p,q)\big)$ if constrained by the condition of *fixed energy*, $H(p,q) = E$. Although essentially equivalent for most applications, the canonical and microcanonical distributions characterize two different situations: systems with and without energy exchange with a heat bath.

For non-interacting particles, $H = \sum_i \big[\boldsymbol{p}_i^2/2m + V(\boldsymbol{q}_i)\big]$, one obtains from (3.20) a factorizing canonical distribution $\rho_\Gamma(p,q) = \prod_i \big[\rho_\mu(\boldsymbol{p}_i,\boldsymbol{q}_i)/N\big]$, as already considered in (3.18), with a μ-space distribution given by $\rho_\mu(\boldsymbol{p},\boldsymbol{q}) \propto N\exp\Big\{-\big[\boldsymbol{p}^2/2m + V(\boldsymbol{q})\big]/kT\Big\}$. This is a Maxwell distribution multiplied by the barometric formula. However, the essential advantage of the canonical Γ-space distribution (3.20) over Boltzmann's is its ability to describe equilibrium correlations between particles. This has been demonstrated in particular by the cluster expansion of Ursell and Mayer (see Mayer and Mayer 1940), in more recent terminology called an expansion by N-point functions, and technically a predecessor of Feynman graphs. However, the distribution (3.20) must not include macroscopic degrees of freedom (such as the position and shape of a solid body). In the case of a rotationally symmetric Hamiltonian, for example, the solid body in thermodynamical equilibrium would otherwise have to be physically characterized by a symmetric distribution of all its orientations in space rather than by a definite orientation. Similarly, its center of mass would always have to be expected close to the minimum of an external potential (see also Fröhlich 1973). These macroscopic variables are *dynamically robust* rather than behaving ergodically. In order to calculate a thermodynamically meaningful representative ensemble according to (3.19), one has to impose additional constraints to fix their values (see Sect. 3.3.1).

Gibbs' extension in phase η thus appears superior to Boltzmann's H-functional (3.9). Unfortunately, the corresponding *ensemble entropy* S_Γ has two (related) defects, which render it entirely unacceptable for representing physical entropy: (1) in stark contrast to the Second Law it remains constant under exact (Hamiltonian) dynamics, and (2) it is obviously not an additive (or extensive) quantity, that would define an entropy *density*.

In order to confirm the first statement, one may formulate the exact ensemble dynamics in Γ-space in analogy to (3.4) by using the $6N$-dimensional continuity equation

$$\frac{\partial \rho_\Gamma}{\partial t} + \mathrm{div}_\Gamma(\rho_\Gamma \boldsymbol{v}_\Gamma) = 0 \; . \tag{3.21}$$

It describes the conservation of probabilities for volume elements moving through Γ-space by forming a bunch of trajectories. The $6N$-dimensional velocity $\boldsymbol{v}_\Gamma$ may be replaced by means of the Hamiltonian equations,

$$\boldsymbol{v}_\Gamma \equiv (\dot{p}_1, \ldots, \dot{p}_{3N}, \dot{q}_1, \ldots, \dot{q}_{3N}) = \left(-\frac{\partial H}{\partial q_1}, \ldots, -\frac{\partial H}{\partial q_{3N}}, \frac{\partial H}{\partial p_1}, \ldots, \frac{\partial H}{\partial p_{3N}} \right) . \tag{3.22}$$

So when rewriting (3.21) by means of the identity

$$\mathrm{div}_\Gamma(\rho_\Gamma \boldsymbol{v}_\Gamma) = \rho_\Gamma \mathrm{div}_\Gamma \boldsymbol{v}_\Gamma + \boldsymbol{v}_\Gamma \cdot \boldsymbol{\nabla}_\Gamma \rho_\Gamma \; ,$$

one may use the *Liouville theorem*,

$$\mathrm{div}_\Gamma \boldsymbol{v}_\Gamma = -\frac{\partial^2 H}{\partial p_1 \partial q_1} - \cdots - \frac{\partial^2 H}{\partial p_{3N} \partial q_{3N}} + \frac{\partial^2 H}{\partial q_1 \partial p_1} + \cdots + \frac{\partial^2 H}{\partial q_{3N} \partial p_{3N}} \equiv 0 \; , \tag{3.23}$$

which describes an incompressible 'fluid' in Γ-space. One thus obtains the *Liouville equation*,

$$\frac{\partial \rho_\Gamma}{\partial t} = -\boldsymbol{v}_\Gamma \cdot \boldsymbol{\nabla}_\Gamma \rho_\Gamma = \sum_{n=1}^{3N} \left(\frac{\partial H}{\partial q_n} \frac{\partial \rho_\Gamma}{\partial p_n} - \frac{\partial H}{\partial p_n} \frac{\partial \rho_\Gamma}{\partial q_n} \right) = \{H, \rho_\Gamma\} \; , \tag{3.24}$$

where $\{a, b\}$ defines the Poisson bracket for two functions a and b. This equation represents the exact Hamiltonian dynamics for ensembles $\rho_\Gamma(p, q, t)$ under the assumption of individually conserved probabilities.

From this analogy with an incompressible fluid in space one may expect the ensemble entropy S_Γ (the measure of 'extension in phase') to remain constant in time. This can indeed be confirmed by differentiating (3.17), inserting (3.24), and repeatedly integrating by parts:

$$\begin{aligned}
\frac{\mathrm{d}S_\Gamma}{\mathrm{d}t} &= \int (\ln\rho_\Gamma + 1)\dot{\rho}_\Gamma \mathrm{d}p\,\mathrm{d}q \\
&= \int (\ln\rho_\Gamma + 1)\sum_{n=1}^{3N}\left(\frac{\partial H}{\partial q_n}\frac{\partial\rho_\Gamma}{\partial p_n} - \frac{\partial H}{\partial p_n}\frac{\partial\rho_\Gamma}{\partial q_n}\right)\mathrm{d}p\,\mathrm{d}q \\
&= -\int \sum_{n=1}^{3N}\left(\frac{\partial H}{\partial q_n}\frac{\partial\ln\rho_\Gamma}{\partial p_n} - \frac{\partial H}{\partial p_n}\frac{\partial\ln\rho_\Gamma}{\partial q_n}\right)\rho_\Gamma\mathrm{d}p\,\mathrm{d}q \\
&= -\int \sum_{n=1}^{3N}\left(\frac{\partial H}{\partial q_n}\frac{\partial\rho_\Gamma}{\partial p_n} - \frac{\partial H}{\partial p_n}\frac{\partial\rho_\Gamma}{\partial q_n}\right)\mathrm{d}p\,\mathrm{d}q = 0\ .
\end{aligned} \tag{3.25}$$

A more instructive proof may be obtained by multiplying the Liouville equation (3.24) by the imaginary unit i in order to cast the dynamics into the *form* of a Schrödinger equation,

$$\mathrm{i}\frac{\partial\rho_\Gamma}{\partial t} = \mathrm{i}\{H,\rho_\Gamma\} =: \hat{L}\rho_\Gamma\ . \tag{3.26}$$

The operator $\hat{L}$ (acting on probability densities) is called the *Liouville operator*. In accordance with this analogy one may use the formal solution $\rho_\Gamma(t) = \exp(-\mathrm{i}\hat{L}t)\rho_\Gamma(0)$, valid if $\partial\hat{L}/\partial t = 0$ (see Prigogine 1962). The Liouville operator is Hermitean with respect to the inner product $\langle\rho_\Gamma,\rho_\Gamma{}'\rangle := \int \rho_\Gamma^*\rho_\Gamma'\mathrm{d}p\,\mathrm{d}q$ (that is, $\langle\rho_\Gamma,\hat{L}\rho_\Gamma'\rangle = \langle\hat{L}\rho_\Gamma,\rho_\Gamma'\rangle$), as can again be shown by partial integration. This means that the Liouville equation conserves these inner products. In particular, for $\rho_\Gamma' = \ln\rho_\Gamma$, one has

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle\rho_\Gamma,\ln\rho_\Gamma\rangle = \frac{\mathrm{d}}{\mathrm{d}t}\overline{\ln\rho_\Gamma} = 0\ , \tag{3.27}$$

since the Liouville operator, when applied to a function $f(\rho_\Gamma)$, satisfies the same Leibniz chain rule $\hat{L}f(\rho_\Gamma) = (\mathrm{d}f/\mathrm{d}\rho_\Gamma)\hat{L}\rho_\Gamma$ as the time derivative.

The *norm* corresponding to this inner product, $\|\rho_\Gamma\|^2 = \langle\rho_\Gamma,\rho_\Gamma\rangle = \int\rho_\Gamma^2\mathrm{d}p\,\mathrm{d}q = \overline{\rho_\Gamma}$, is then also dynamically invariant. It represents a linear measure of extension in phase (a *linear ensemble entropy*[3]), and thus has to be distinguished from the probability norm $\int\rho_\Gamma\mathrm{d}p\,\mathrm{d}q = \overline{1} = 1$. The conservation of these measures under a Liouville equation confirms in turn that the Γ-space volume is an appropriate measure for non-countable sets of states (Ehrenfest and Ehrenfest 1911): the thus defined 'number' of states does not change under an appropriately defined determinism. A more fundamental justification of this measure can be derived from the conservation of probabilities of discrete *quantum* states (see Sect. 4.1).

[3] See Wehrl (1978) for further measures, which are, however, not always monotonically related to one another. The conventional logarithmic measure is usually preferred because of the resulting additivity of the entropies of statistically independent subsystems.

The conservation of ensemble entropy, implied by using the exact dynamics, is unacceptable in a statistical foundation of *physical* entropy. Therefore, Gibbs introduced a more subtle concept of entropy, that was motivated by his famous *ink drop analogy*: A bit of ink dropped into a glass of water is assumed to behave as an incompressible fluid when the water is stirred. Although its volume must remain constant, the whole glass of water will soon appear homogeneous in light blue. Only a microscopic examination would reveal that the ink had simply rearranged itself into many thin tubes, which are everywhere dense in spite of occupying only a volume of the initial size of the droplet.

Therefore, Gibbs defined his new entropy S_{Gibbs} by means of a *coarse-grained* distribution ρ^{cg}, obtained by averaging over small (but fixed) $6N$-dimensional volume elements ΔV_m $(m = 1, 2, \ldots)$ which cover the whole $\mathit{\Gamma}$-space:

$$\rho^{\mathrm{cg}}(p,q) = \frac{1}{\Delta V_m}\int_{\Delta V_m} \rho(p',q')\mathrm{d}p'\mathrm{d}q' =: \frac{\Delta p_m}{\Delta V_m}\ , \quad \text{for } p,q \in \Delta V_m\ . \qquad (3.28)$$

The resulting ensemble entropy is then given by

$$S_{\mathrm{Gibbs}} := -k\eta[\rho^{\mathrm{cg}}] = -k\sum_m \Delta p_m \ln\frac{\Delta p_m}{\Delta V_m}\ . \qquad (3.29)$$

As already mentioned in connection with the smoothing of Boltzmann's μ-space distributions, the justification of this procedure by means of the quantum mechanical uncertainty relations, that is, by coarse-graining over phase space cells of size h^{3N}, may be tempting, but would clearly be inconsistent with classical mechanics. The consistent quantum mechanical treatment (Chap. 4) leads again to the conservation of ensemble entropy (now for ensembles of wave functions rather than $\mathit{\Gamma}$-space points). 'Quantum cells' of size h^{3N} can be justified only as convenient *units* of phase space volume in order to obtain the same normalization of entropy as in the classical limit of *quantum* statistical mechanics, where ensemble entropy vanishes for pure states, which correspond to phase space 'cells' – see (4.21). However, these quantum cells do *not* define uncertain initial conditions which might explain quantum indeterminism (as often claimed); ensemble entropy is conserved under Hamiltonian *and* Schrödinger dynamics.

The increase in Gibbs' entropy can be understood according to the classical ink drop analogy. While the volume of the compact ink droplet is only slightly increased by moderate coarse-graining, that of a dense web of thin tubes (obtained by stirring) is considerably enlarged. Even though the coarse-graining itself is artificial, its efficiency depends on the shape of the volume to which it is applied. This is similar to using Boltzmann's smooth μ-space densities, which characterize properties of the *discrete* particle distributions. Since there evidently exist far more droplet shapes with a large surface than compact ones, the former have to be regarded as more probable. For *statistical* reasons one should hardly ever find a compact droplet (which is confirmed for three-dimensional 'droplets' in the absence of any surface tension).

However, there remains an essential difference between a droplet of ink in water and a dynamical volume element in phase space. While the extension and shape of a droplet are real physical properties, the real state of a classical mechanical system is represented by a *point* in phase space. Coarse-graining of the ink drop may be likened to Boltzmann's smoothing procedure in-so-far as it preserves properties of the discrete particle distribution, while Gibbs' entropy for a real state p, q, $S_{\text{Gibbs}} = f(p,q) := k \ln \Delta V_{m_0}$ (resulting if $p, q \in \Delta V_{m_0}$), would be entirely artificial. This difference would be reduced if individual classical state were identified with $N!$ points in $\varGamma-$space.

Gibbs' procedure is therefore usually applied to presumed phase space densities, which can only represent incomplete *information*. His entropy then measures the enlargeability by coarse-graining of a certain state of knowledge – not by coarse-graining of a real physical state. Its increase, $\mathrm{d}S_{\text{Gibbs}}/\mathrm{d}t \geq 0$, under a deterministic (information-conserving) dynamical law describes the transformation of macroscopic information, assumed to be present *initially*, into fine-grained information, that is then regarded as 'irrelevant' and dynamically neglected (Sect. 3.2). However, this procedure may be in conflict with the idea of entropy as an objective physical quantity that is independent of any information held by an observer. This fundamental problem will be addressed again in Sect. 3.3 and later chapters.

Similar to the problem that arose for μ-space densities in (3.2), the coarse-graining cannot be uniquely inverted, since it destroys information. The intended chain of calculation,

$$\rho^{\text{cg}} \longrightarrow \rho \quad \xrightarrow{\hat{L}} \quad \frac{\partial \rho}{\partial t} \longrightarrow \frac{\partial \rho^{\text{cg}}}{\partial t} \,, \tag{3.30}$$

is again broken at its first link. A new autonomous dynamics has therefore been proposed for ρ^{cg}, in analogy to the *Stoßzahlansatz*, by complementing the Hamiltonian dynamics with a *dynamical* coarse-graining, applied in small but finite time steps Δt:

$$\left\{ \frac{\partial \rho^{\text{cg}}}{\partial t} \right\}_{\text{master}} := \frac{\left[\mathrm{e}^{-\mathrm{i}\hat{L}\Delta t} \rho^{\text{cg}} \right]^{\text{cg}} - \rho^{\text{cg}}}{\Delta t} \,. \tag{3.31}$$

In this form it may also be regarded as a variant of a 'unifying principle' that was proposed as a stochastic process by R.M. Lewis (1967). Instead of dynamically applying Gibbs' coarse-graining in (3.31), Lewis suggested *maximizing the entropy* in each dynamical step under the constraint of certain fixed 'macroscopic' quantities (see also Jaynes' theory in Sect. 3.3.1).

Equation (3.31) defines reasonable dynamics if the corresponding probability increments $\Delta\Delta p_m$ – see (3.28) – are proportional to Δt for small but finite time intervals Δt, thus describing transition *rates* between the cells ΔV_m. This important condition will be discussed in a more general form in Sect. 3.2, and later for deriving the Pauli equation (4.18). Master equations such as (3.31) ensure a monotonic entropy increase. Their approximate validity requires that

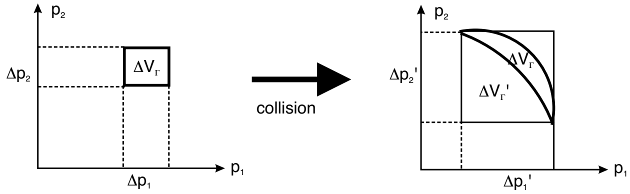

Fig. 3.1. Transformation of information about particle momenta into information about correlations between them as the basis of the H-theorem (symbolic, for non-central collisions)

the arising microscopic (fine-grained) information remains *dynamically* irrelevant for the evolution of the coarse-grained distribution. Except in the case of equilibrium, this cannot simultaneously be true in different directions of time.

The meaning of Boltzmann's *Stoßzahlansatz* (3.6) can be similarly understood, as it neglects all particle correlations, which are thus regarded as fine-grained information, *after* they have formed in collisions. It is again based on the assumption that the interval Δt is finite and large compared to collision times. The effect of an individual collision on the phase space distribution may be illustrated in two-dimensional momentum space (Fig. 3.1): a collision between two particles with small momentum uncertainties Δp_1 and Δp_2 leads deterministically to a *correlating* (deformed) volume element of the same size ΔV_Γ as the initial one. (In a realistic description, momenta would also be correlated with particle positions.) Subsequent neglect of the arising correlations will then enlarge this volume element ($\Delta V'_\Gamma > \Delta V_\Gamma$). However, neglecting such statistical correlations evidently has no effect on a phase space *point*.

The question as to the precise mathematical conditions under which certain systems are indeed 'mixing' in the sense of the plausible ink drop analogy (in a stronger version referred to as *K-systems* after Kolmogorov) is rigorously investigated, though under idealized conditions (such as ideal isolation), in *ergodic theory* (see Arnol'd and Avez 1968, or Mackey 1989). Most non-ergodic systems are pathological in forming sets of measure zero, or in being unstable against unavoidable perturbations. In general, the quantitative question for the *time-scale* of mixing between different regions in phase space is physically far more relevant than exact formal theorems which apply only at infinite times. Regions which don't mix with others over long times may define robust (usually macroscopic) properties, that do *not* have to represent constants of the motion. On the other hand, certain non-ergodic aspects have been claimed to apply under quite general circumstances (Yoccoz 1992), but no physically relevant interpretation of these formal dynamical properties has ever been given.

The strongest mixing is required for the finest conceivable coarse-graining. This is given by its nontrivial limit $\Delta V_\Gamma \to 0$ for the size of grains, which defines a *weak convergence* for measures on phase space. It would again lead to infinite Poincaré recurrence times for isolated systems. However, this is neither required in a universe of finite age, nor would it be realistic, since quantum theory limits the entropy capacity available in the form of unlimited fine-graining of classical phase space. For this quantum mechanical reason there can be no 'overdetermination' of the *microscopic* past in spite of the validity of microscopic causality (see footnote 1 of Chap. 2 and the end of Sect. 5.3). However, it is important to note that all concepts of mixing are T-symmetric. In order to explain the time asymmetry of the Second Law ('irreversibility'), they would have to be *applied dynamically* in a specific direction of time.

Dynamical coarse-graining as in (3.31) may also be based on an incompletely known Hamiltonian. An *ensemble of Hamiltonians* defines a stochastic dynamical model when used for calculating 'forward' in time. Even very small uncertainties in the Hamiltonian may be sufficient to completely destroy fine-grained information within a short time interval. Borel (1924) estimated the effect of a gravitational force that would arise here on earth by the displacement of a mass of the order of a few grams by a few centimeters at the distance of Sirius. He thereby pointed out that this would lead to a completely different microscopic state for the molecules forming a gas in a vessel under normal conditions within seconds. Although distortions of the individual molecular trajectories are extremely small, they would be amplified in each subsequent collision by a factor of the order of l/R, the ratio of the mean free path over the molecular radius. This extreme sensitivity to the environment describes in effect a *local microscopic indeterminism*.[4] In many situations, the microscopic distortions may even co-determine macroscopic effects (thus inducing an effective *macroscopic indeterminism*), as discussed, in particular, in the *theory of chaos* ('butterfly effect').

The essence of Borel's argument is that macroscopic systems, aside from the whole Universe, may never be regarded as dynamically isolated – even when thermodynamically closed in the sense of $\mathrm{d}S_{\mathrm{ext}} = 0$. The *dynamical* coarse-graining that is part of the master equation (3.31) may indeed be ascribed to perturbations by the environment – provided the latter obey causality, that is, can be treated stochastically in the forward direction of time. This important dynamical assumption is yet another form of the *intuitive causality* discussed at the beginning of Chap. 2 as a major manifestation of the arrow of time. The *representative ensembles* used in statistical thermodynamics may therefore be understood within classical physics as those which arise (and are maintained) by this stochastic nature of unavoidable perturbations, while 'robust' properties can be regarded as macroscopic.

While the *intrinsic dynamics* of a macroscopic physical system transforms coarse-grained into fine-grained information, interactions with the environment thus transform the resulting fine-grained information very efficiently into practically useless correlations with distant systems. The sensitivity of

---

[4] While the effect of Borel's gravitational distortion is drastically reduced for quantized interactions, other environmental effects (such as decoherence) then become important in producing an effective local indeterminism (see Sects. 4.3.4 and 5.3).

the microscopic states of macroscopic systems to such interactions with their environments strongly indicates that simultaneously existing *opposite arrows of time* in different regions of the Universe would be inconsistent with one another. This universality of the arrow of time seems to be its most important property. Time asymmetry has therefore been regarded as a global *symmetry breaking*. However, such a conclusion would *not* exclude the far more probable situation of thermal equilibrium.

Lawrence Schulman (1999) has challenged the usual assumption of a universal arrow of time by suggesting explicit counterexamples. Most of them are indeed quite illustrative in emphasizing the role of initial of final conditions, but they appear unrealistic in our Universe (see Zeh 2005b). The situation is similar to the symmetric boundary conditions suggested by Wheeler and Feynman in electrodynamics, and discussed in Sect. 2.4. Local final conditions at the present stage of the Universe or in the near future can hardly be retrocaused by a low entropy condition at the big crunch (see also Casati, Chirikov and Zhirov 2000), but may be essential during a conceivable recontraction era of the Universe (see Sect. 5.3).

In order to reverse the thermodynamical arrow of time in a bounded system, it would not therefore suffice to "go ahead and reverse all momenta" in the system itself, as ironically suggested by Boltzmann as an answer to Loschmidt. In an interacting Laplacean universe, the Poincaré cycles of its subsystems could in general only be those of the whole Universe, since their exact Hamiltonians must always depend on their time-dependent environment.

Time reversal including thermodynamical aspects has been achieved even in practice for very weakly interacting spin waves (Rhim, Pines and Waugh 1971). The latter can be regarded as isolated systems to a very good approximation (similar to electromagnetic waves in the absence of absorbers), while allowing a sudden sign reversal of their spinor Hamiltonian in order to simulate time reversal ($dt \to -dt$). These spin wave experiments demonstrate that a closed system in thermodynamical equilibrium may preserve an arrow of time in the form of *hidden correlations*. When a closed system has reached macroscopic equilibrium, it *appears* T-symmetric, although its fine-grained information determines the distance and direction in time to its low-entropy state in the past (see also the Appendix for a numerical example). In contrast to such rare almost-closed systems, generic ones are strongly affected by Borel's argument, and *cannot* be reversed by local manipulations.

3.2 Zwanzig's General Formalism of Master Equations

Boltzmann's *Stoßzahlansatz* (3.6) for μ-space distributions and the master equation (3.31) for coarse-grained Γ-space distributions can thus be understood in a similar way. They describe the transformation of *special* macroscopic states into more probable ones, whereby the higher information content of the former is transformed into macroscopically irrelevant information. There are many other master equations based on the same strategy, and designed to suit various purposes. Zwanzig (1960) succeeded in formalizing them in a general and instructive manner that also reveals their analogy with retarded electrodynamics as another manifestation of the arrow of time – see (3.40)–(3.49) below.

The basic concept of Zwanzig's formalism is defined by idempotent mappings $\hat{P}$, acting on probability distributions $\rho(p,q)$:

$$\rho \;\rightarrow\; \rho_{\mathrm{rel}} := \hat{P}\rho\,, \quad \text{with} \quad \hat{P}^2 = \hat{P} \quad \text{and} \quad \rho_{\mathrm{irrel}} := (1-\hat{P})\rho\,. \tag{3.32}$$

Their meaning will be illustrated by means of several examples below, before explaining the dynamical formalism. If these mappings *reduce* the information content of ρ to what is then called its 'relevant' part ρ_{rel}, they may be regarded as a *generalized coarse-graining*. In order to interpret ρ_{rel} as a probability density again, one has to require its non-negativity and, for convenience,

$$\int \rho_{\mathrm{rel}} \mathrm{d}p\,\mathrm{d}q = \int \rho \mathrm{d}p\,\mathrm{d}q = 1\,, \tag{3.33}$$

that is,

$$\int \rho_{\mathrm{irrel}} \mathrm{d}p\,\mathrm{d}q = \int (1-\hat{P})\rho \mathrm{d}p\,\mathrm{d}q = 0\,. \tag{3.34}$$

Reduction of information means

$$S_\Gamma[\hat{P}\rho] \geq S_\Gamma[\rho] \tag{3.35}$$

(or similarly for any other measure of ensemble entropy).

Using this concept, Lewis' master equation (3.31), for example, may be written in the generalized form

$$\left\{\frac{\partial \rho_{\mathrm{rel}}}{\partial t}\right\}_{\mathrm{master}} := \frac{\hat{P}\mathrm{e}^{-\mathrm{i}\hat{L}\Delta t}\rho_{\mathrm{rel}} - \rho_{\mathrm{rel}}}{\Delta t}\,. \tag{3.36}$$

It would then describe a monotonic increase in the corresponding entropy $S[\rho_{\mathrm{rel}}]$. In contrast to Zwanzig's approach, to be described below, phenomenological master equations such as Lewis's unifying principle have often been meant to describe a *fundamental* indeterminism that would replace reversible Laplacean determinism.

In most applications, Zwanzig's idempotent operations $\hat{P}$ are linear and Hermitean with respect to the inner product for probability distributions defined above (3.27). In this case they are projection operators, which preserve only some 'relevant component' of the original information. If such a projection obeys (3.33) for every ρ, it must leave the equipartition invariant, $\hat{P}1 = 1$, as can be shown by writing down the above-mentioned inner product of this equation with an *arbitrary* distribution ρ and using the hermiticity of $\hat{P}$.

Zwanzig's dynamical formalism may also be useful for non-Hermitean or even non-linear idempotent mappings $\hat{P}$ (see Lewis 1967, Willis and Picard 1974). These mappings are then not projections any more: they may even *create new information*. A trivial example for the creation of information is the nonlinear mapping of all probability distributions onto a fixed one, $\hat{P}\rho := \rho_0$ for all ρ, regardless of whether or not they contain a component proportional to ρ_0. The physical meaning of such generalizations of Zwanzig's formalism will be discussed in Sects. 3.4 and 4.4. In the following we shall consider information-reducing mappings.

Zwanzig's 'projection' concept is deliberately kept general in order to permit a wealth of applications. Examples introduced so far are coarse-graining, $\hat{P}_{\mathrm{cg}}\rho := \rho^{\mathrm{cg}}$, as defined in (3.28), and the neglect of correlations between particles by means of μ-space densities:

$$\hat{P}_\mu \rho(p,q) := \prod_{i=1}^{N} \frac{\rho_\mu(\boldsymbol{p}_i, \boldsymbol{q}_i)}{N} \ ,$$

with

$$\rho_\mu(\boldsymbol{p},\boldsymbol{q}) := \sum_{i=1}^{N} \int \rho(p,q)\delta^3(\boldsymbol{p}-\boldsymbol{p}_i)\delta^3(\boldsymbol{q}-\boldsymbol{q}_i)\mathrm{d}p\,\mathrm{d}q \ . \tag{3.37}$$

(As before, boldface letters represent three-dimensional vectors, while p, q is a point in Γ-space.) The latter example defines a *non-linear* though information-reducing 'Zwanzig projection'. Most arguments applying to linear operators $\hat{P}$ remain valid in this case when applied to the linearly resulting μ-space distributions $\rho_\mu(\boldsymbol{p},\boldsymbol{q})$ (which do not live in Γ-space) rather than to their products $\hat{P}_\mu\rho(p,q)$ (which do). In quantum theory, this approach is related to the Hartree or mean field approximation. Boltzmann's 'relevance concept', which, when written as a Zwanzig projection, would map real states onto products of *smooth* μ-space distributions, can then be written as $\hat{P}_{\mathrm{Boltzmann}} = \hat{P}_\mu \hat{P}_{\mathrm{cg}}$. An obvious generalization of $\hat{P}_{\mathrm{Boltzmann}}$ can be defined by a projection onto two-particle correlation functions. In this way, a complete hierarchy of relevance concepts in terms of *n-point functions* (equivalent to a cluster expansion) can be defined.

A particularly important concept of relevance, that is often not even noticed, is *locality* (see, e.g., Penrose and Percival 1962). It is required in order to define entropy as an extensive quantity – in accordance with the phenomenological equation (3.1) and with the concept of an entropy *density* $s(\boldsymbol{r})$, such that $S = \int s(\boldsymbol{r})\mathrm{d}^3 r$. The corresponding Zwanzig projection of locality may be symbolically written as

$$\hat{P}_{\mathrm{local}}\rho := \prod_k \rho_{\Delta V_k} \ . \tag{3.38}$$

The RHS here is meant to describe the neglect of all statistical correlations beyond a distance defined by the size of volume elements ΔV_k. The probability distributions $\rho_{\Delta V_k}$ would here be defined by integrating over all external degrees of freedom. The volume elements have to be chosen large enough to contain a sufficient number of particles in order to preserve dynamically relevant short range correlations (as required for real gases, for example). In order to allow volume elements ΔV_k with physically open boundaries, their probability distributions $\rho_{\Delta V_k}$ in (3.38) have to admit variable particle number (density fluctuations) – as in a grand canonical ensemble.

Locality is presumed, in particular, when writing (3.1) in its differential (local) form as a 'continuity inequality' for the entropy density $s(\boldsymbol{r}, t)$,

$$\frac{\partial s}{\partial t} + \mathrm{div}\boldsymbol{j}_s \geq 0 \, , \tag{3.39}$$

with an entropy current density $\boldsymbol{j}_s(\boldsymbol{r}, t)$. This form allows the definition of phenomenological entropy-producing (hence positive) terms on the RHS in order to replace the inequality by an *equation* (see Landau and Lifschitz 1959 or Glansdorff and Prigogine 1971). An example is the source term $\kappa(\nabla T)^2/T^2$ in the case of heat conduction, where κ is the heat conductivity.

The general applicability of (3.39) demonstrates that the concept of *physical entropy* is always based on the neglect of nonlocal correlations. Therefore, the production of entropy can be usually understood as the transformation of local information into nonlocal correlations (as depicted in Fig. 3.1). This description is in accordance with the conservation of ensemble entropy (determinism) *and* with intuitive causality. The Second Law thus depends crucially on the dynamical irrelevance of microscopic correlations *for the future* (as assumed in the *Stoßzahlansatz*, for example). Since this 'microscopic causality' cannot be observed as easily and directly as the causal correlations which define retardation of macroscopic radiation, its validity under all circumstances has been questioned (Price 1996). However, it is not only indirectly confirmed by the success of the *Stoßzahlansatz*, but also (in its quantum mechanical form – see Sect. 4.2) by the validity of a Sommerfeld radiation condition (see Sect. 2.1) for microscopic scattering experiments, or by the validity of exponential decay (Sect. 4.5).

The Zwanzig projection of locality is again ineffective on real states, which are always local in the sense of defining the states of all their subsystems. Therefore, applying $\hat{P}_{local}$ to an individual state (a δ-function or sum of them) would not lead to a non-singular entropy S_Γ. This will drastically change in quantum mechanics, because it is kinematically non-local (Chap. 4).

As already mentioned on p. 55, coarse-graining as a relevance concept may also enter in a hidden form, corresponding to its nontrivial limit $\Delta V_\Gamma \to 0$, by considering only *non-singular measures* on phase space (thus excluding δ-functions). This strong idealization may be mathematically signalled by the 'unitary inequivalence' of the original Liouville equation and the master equations resulting in this limit (see Misra 1978 or Mackey 1989).

Further examples of Zwanzig projections will be defined throughout the book, in particular in Chap. 4 for quantum mechanical applications, where the relevance of locality leads to the important concept of decoherence. Different schools and methods of irreversible thermodynamics may even be distinguished according to the concepts of relevance which they are using, and which they typically regard as 'natural' or 'fundamental' (see Grad 1961).

However, the mere *conceptual* foundation of a relevance concept ('paying attention' only to certain aspects) is insufficient for justifying its *dynamical* autonomy in the form of a master equation (3.36) – see the Appendix for an explicit example. Locality *is* usually dynamically relevant in this sense because of the locality of all interactions. This dynamical locality is essential even for the very concept of physical *systems*, including those of local observers as the ultimate referees for what is relevant.

Zwanzig reformulated the exact Hamiltonian dynamics for $\rho_{\rm rel}$ regardless of any specific choice of $\hat{P}$ instead of simply *postulating* a phenomenological master equation (3.36) in analogy to Boltzmann or Lewis. It can then in general not be autonomous[5], that is, of the form $\partial\rho_{\rm rel}/\partial t = f(\rho_{\rm rel})$, but has to be written as

$$\frac{\partial\rho_{\rm rel}}{\partial t} = f(\rho_{\rm rel}, \rho_{\rm irrel}) \tag{3.40}$$

in order to eliminate $\rho_{\rm irrel}$ by means of certain assumptions. The procedure is analogous to the elimination of the electromagnetic degrees of freedom by means of the condition $A^\mu_{\rm in} = 0$ when deriving a retarded action-at-a-distance theory (Sect. 2.2). In both cases, empirically justified boundary conditions which specify a time direction are assumed to hold for the degrees of freedom that are to be eliminated.

To this end the Liouville equation $\mathrm{i}\partial\rho/\partial t = \hat{L}\rho$ is decomposed into its relevant and irrelevant parts by multiplying it by $\hat{P}$ or $1-\hat{P}$, respectively:

$$\mathrm{i}\frac{\partial\rho_{\rm rel}}{\partial t} = \hat{P}\hat{L}\rho_{\rm rel} + \hat{P}\hat{L}\rho_{\rm irrel} \; , \tag{3.41a}$$

$$\mathrm{i}\frac{\partial\rho_{\rm irrel}}{\partial t} = (1-\hat{P})\hat{L}\rho_{\rm rel} + (1-\hat{P})\hat{L}\rho_{\rm irrel} \; . \tag{3.41b}$$

This corresponds to representing the Liouville operator by a matrix of operators

$$\hat{L} = \begin{pmatrix} \hat{P}\hat{L}\hat{P} & \hat{P}\hat{L}(1-\hat{P}) \\ (1-\hat{P})\hat{L}\hat{P} & (1-\hat{P})\hat{L}(1-\hat{P}) \end{pmatrix} \; . \tag{3.42}$$

Equation (3.41b) for $\rho_{\rm irrel}$, with $(1-\hat{P})\hat{L}\rho_{\rm rel}$ regarded as an inhomogeneity, may then be formally solved by the method of the variation of constants (interaction representation). This leads to

$$\rho_{\rm irrel}(t) = \mathrm{e}^{-\mathrm{i}(1-\hat{P})\hat{L}(t-t_0)}\rho_{\rm irrel}(t_0) - \mathrm{i}\int_0^{t-t_0} \mathrm{e}^{-\mathrm{i}(1-\hat{P})\hat{L}\tau}(1-\hat{P})\hat{L}\rho_{\rm rel}(t-\tau)\mathrm{d}\tau \; , \tag{3.43}$$

[5] In mathematical physics, 'autonomous dynamics' is often defined as the absence of any explicit time dependence in the dynamics – regardless of whether it is fundamental or caused by a time-dependent environment.

as may be confirmed by differentiation.

If $t > t_0$, (3.43) is analogous to the *retarded form* (2.9) of the boundary value problem in electrodynamics. In this case, $\tau \geq 0$, and $\rho_{\rm rel}(t-\tau)$ may be interpreted as an advanced source for the 'retarded' $\rho_{\rm irrel}(t)$. Substituting this formal solution (3.43) into (3.41a) leads to three terms on the RHS, viz.,

$$\mathrm{i}\frac{\partial \rho_{\rm rel}(t)}{\partial t} = \mathrm{I} + \mathrm{II} + \mathrm{III} \qquad (3.44)$$
$$\equiv \hat{P}\hat{L}\rho_{\rm rel}(t) + \hat{P}\hat{L}\mathrm{e}^{-\mathrm{i}(1-\hat{P})\hat{L}(t-t_0)}\rho_{\rm irrel}(t_0) - \mathrm{i}\int_0^{t-t_0} \hat{G}(\tau)\rho_{\rm rel}(t-\tau)\mathrm{d}\tau \, .$$

The integral kernel of the last term,

$$\hat{G}(\tau) := \hat{P}\hat{L}\mathrm{e}^{-\mathrm{i}(1-\hat{P})\hat{L}\tau}(1-\hat{P})\hat{L}\hat{P} \, , \qquad (3.45)$$

corresponds to the retarded Green's function of Sect. 2.1.

Equation (3.44) is exact and, therefore, cannot yet describe time asymmetric dynamics. Since it forms the first step in this derivation of master equations, it is known as a *pre-master equation*. The meanings of its three terms are illustrated in Fig. 3.2. The first one describes the internal dynamics of $\rho_{\rm rel}$. In Boltzmann's μ-space dynamics (3.3), it would correspond to $\{\partial\rho_\mu/\partial t\}_{\rm free+ext}$. It vanishes if $\hat{P}\hat{L}\hat{P} = 0$ (as is often the case).[6]

The second term of (3.44) is usually omitted by presuming the absence of irrelevant initial information: $\rho_{\rm irrel}(t_0) = 0$. If *relevant* information happens to be present initially, it can then be dynamically transformed into irrelevant information. (Because of the asymmetry between $\hat{P}$ and $1-\hat{P}$, irrelevant information would have to be measured by $-S_\Gamma[\rho] + S_\Gamma[\rho_{\rm rel}]$ rather than by $-S_\Gamma[\rho_{\rm irrel}]$.)

The vital third term is *non-Markovian* (non-local in time), as it depends on the whole time interval between t_0 and t. Its retarded form (valid for $t > t_0$) is compatible with the intuitive concept of causality. This term becomes approximately *Markovian* if $\rho_{\rm rel}(t-\tau)$ varies slowly for a small 'relaxation time' τ_0 during which $\hat{G}(\tau)$ becomes negligible for reasons to be discussed. In (3.44), $\hat{G}(\tau)$ may then be regarded to lowest order as being proportional to a δ-function in τ. This assumption is also contained in Boltzmann's *Stoßzahlansatz*, where it means that correlations arising by scattering

[6] Since the (indirectly acting) non-trivial terms contribute only in second and higher orders of time, the time derivative defined by the master equation (3.36) would then vanish in the limit $\Delta t \to 0$. This corresponds to what in quantum theory is known as the *quantum Zeno paradox* (Misra and Sudarshan 1977), also called *watched pot behavior* or the *watchdog effect*. It describes an *immediate* loss of information from the irrelevant channel (or its dynamically relevant parts – see later in the discussion), such that it has no chance of affecting its relevant counterpart any more. Fast information loss may be caused by a strong coupling to the environment, for example. Since this efficiency depends on the energy level density (Joos 1984), the Zeno effect is relevant mainly in quantum theory.

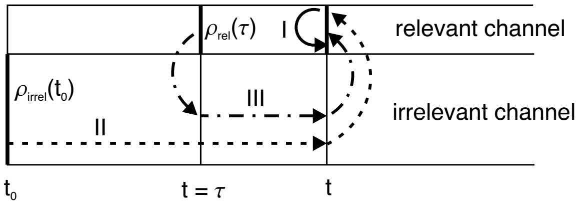

Fig. 3.2. Retarded form of the exact dynamics for the relevant information according to Zwanzig's *pre-master equation* (3.24). In addition to the instantaneous direct interaction I, there is the contribution II arising from the 'incoming' irrelevant information, and the retarded term III in analogy to electromagnetic action at a distance, resulting from 'advanced sources' in the whole time interval between t_0 and t (cf. the left part of Fig. 2.2)

are irrelevant for the forward dynamics of ρ_{rel}. In analogy to retarded electromagnetic forces, this third term of the pre-master equation then assumes the form of an effective direct interaction between the relevant degrees of freedom (though instantaneous in this nonrelativistic treatment). In electrodynamics, the charged sources would represent the 'relevant' variables, while their effective interactions act 'at a distance'. In statistical physics, this 'interaction' describes the dynamics of *ensembles*.

The Markovian approximation may be understood by means of assumptions which simultaneously explain the applicability of the initial condition $\rho_{\mathrm{irrel}} \approx 0$ at *all* times – provided it holds in an appropriate form in the very distant past. This is again analogous to the condition in electrodynamics that A^{μ}_{in} either vanishes or can be well understood in terms of a limited number of known or at least plausible sources *at all times*.

Consider the action of the operator $(1 - \hat{P})\hat{L}\hat{P}$ appearing on the RHS of the kernel (3.45). Because of the structure of a typical Liouville operator, it transforms information from ρ_{rel} only into *specific* parts of ρ_{irrel}. In the scattering theory of complex objects, similar formal parts are called *doorway states* (Feshbach 1962). For example, if the Hamiltonian contains no more than two-particle interactions, $\hat{L}\hat{P}_{\mu}$ creates two-particle correlations. Only the subsequent application of the propagator $\exp[-\mathrm{i}(1 - \hat{P})\hat{L}\tau]$ is then able to produce states 'deeper' in the irrelevant channel (many-particle correlations in this case) – see Fig. 3.3. Recurrence from the depth of the irrelevant channel is related to Poincaré recurrence times, and may in general be neglected (as exemplified by the success of Boltzmann's collision equation). If the *relaxation time*, now defined as the time required for the transfer of information from the doorway 'states' into deeper parts of the irrelevant channel, is of the order τ_0, say, one may assume $\hat{G}(\tau) \approx 0$ for $\tau \gg \tau_0$, as required for the Markovian δ-function approximation $\hat{G}(\tau) \approx \hat{G}_0\delta(\tau)$.

Essential for the validity of this approximation is the large information capacity of the irrelevant channel (similar to that of the electromagnetic field

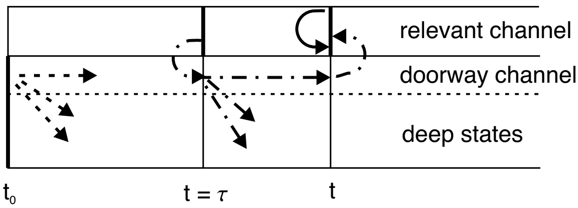

Fig. 3.3. The large information capacity of the irrelevant channel and the specific structure of the interaction together enforce the disappearance of information into the depth of the irrelevant channel if an appropriate initial condition holds

in Chap. 2, but far exceeding it). For example, correlations between particles may describe far more information than the single-particle distribution ρ_μ. A fundamental *cosmological* assumption,

$$\rho_{\rm irrel}(t_0) = 0 \ , \tag{3.46}$$

at a time t_0 in the *finite* past (similar to the cosmological $A^\mu_{\rm in} = 0$ at the big bang) is therefore quite powerful – even though it is a *probable* condition. Any irrelevant information formed later from the initial 'information' contained in $\rho_{\rm rel}(t_0)$ (that is, from any specification of the initial state) may be expected to remain dynamically negligible in (3.44) for a very long time. It would be essential, however, for calculating backwards in time under these conditions.

The assumption $\rho_{\rm irrel} \approx 0$ has thus to be understood in a *dynamical* sense: any newly formed contribution to $\rho_{\rm irrel}$ must remain irrelevant in the 'forward' direction of time. The dynamics for $\rho_{\rm rel}$ may then appear autonomous (while it cannot be exact). For example, all correlations between subsystems seem to require advanced local *causes*, but no similar (retarded) *effects*. Otherwise they would be interpreted as a *conspiracy*, the deterministic version of *causae finales*.

Under these *assumptions*, one obtains from (3.44), as a first step, the non-Markovian dynamics

$$\frac{\partial \rho_{\rm rel}(t)}{\partial t} = - \int_0^{t-t_0} \hat{G}(\tau)\rho_{\rm rel}(t-\tau){\rm d}\tau \ . \tag{3.47}$$

The upper boundary of the integral can here be replaced by a constant T that is large compared to τ_0, but small compared to any (theoretical) recurrence time for $\hat{G}(\tau)$. If $\rho_{\rm rel}(t)$ is now assumed to remain constant over time intervals of the order of the relaxation time τ_0, corresponding to an already prevailing partial (e.g., local) equilibrium, one obtains the time-asymmetric Markovian limit:

$$\frac{\partial \rho_{\rm rel}(t)}{\partial t} \approx -\hat{G}_{\rm ret}\rho_{\rm rel}(t) \ , \tag{3.48}$$

with

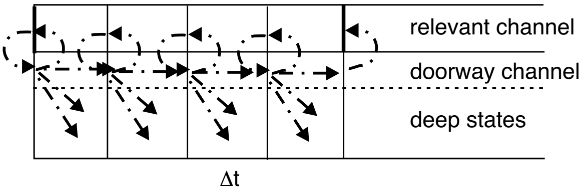

Fig. 3.4. The master equation represents 'alternating dynamics', usually describing a monotonic loss of relevant information

$$\hat{G}_{\mathrm{ret}} := \int_0^T \hat{G}(\tau)\mathrm{d}\tau \; . \tag{3.49}$$

A similar nontrivial limit of vanishing retardation ($\tau_0 \to +0$) led to the LAD equation with its asymmetric radiation reaction in Sect. 2.3. The integral (3.49) could be formally evaluated when inserting (3.45), but it is usually more conveniently computed after this operator has been applied to a specific $\rho(t)$. (See the explicit evaluation for discrete quantum mechanical states in Sect. 4.1.2.)

The autonomous master equation (3.48) again describes *alternating dynamics* of the type (3.36) (see Fig. 3.4). Irrelevant information is disregarded after short time intervals Δt (now representing the relaxation time τ_0). If $\hat{P}$ only *destroys* information, the master equation describes never-decreasing entropy:

$$\frac{\mathrm{d}S_\Gamma[\rho_{\mathrm{rel}}]}{\mathrm{d}t} \geq 0 \; . \tag{3.50}$$

This corresponds to a *positive* operator $\hat{G}_{\mathrm{ret}}$ (as can most easily be shown by means of the *linear* measure of entropy).

A phenomenological probability-conserving Markovian master equation for a system with 'macroscopic states' described by a (set of) 'relevant' variable(s) α, that is, $\rho_{\mathrm{rel}}(t) \equiv \rho(\alpha, t)$ (see also Sects. 3.3 and 3.4) can be written in the general form

$$\frac{\partial \rho(\alpha, t)}{\partial t} = \int \Big[w(\alpha, \alpha')\rho(\alpha', t) - w(\alpha', \alpha)\rho(\alpha, t) \Big] \mathrm{d}\alpha' \; . \tag{3.51}$$

The transition rates $w(\alpha, \alpha')$ here define the phenomenological operator $\hat{G}_{\mathrm{ret}}$ by means of its integral kernel $\hat{G}_{\mathrm{ret}}(\alpha, \alpha') = -w(\alpha, \alpha') + \delta(\alpha, \alpha') \int w(\alpha, \alpha'')\mathrm{d}\alpha''$. They often satisfy a generalized time inversion symmetry,

$$\frac{w(\alpha, \alpha')}{\sigma(\alpha)} = \frac{w(\alpha', \alpha)}{\sigma(\alpha')} \; , \tag{3.52}$$

where $\sigma(\alpha)$ may represent the density of the ('irrelevant') microscopic states with respect to the variable α – that is, $\sigma := \mathrm{d}n/\mathrm{d}\alpha$, where n is the number of microscopic states as a function of α. In this case one may again derive an H-theorem, in analogy to (3.12), for the *generalized H-functional*

$$H_{\text{gen}}[\rho(\alpha)] := \int \rho(\alpha) \ln \frac{\rho(\alpha)}{\sigma(\alpha)} d\alpha = \overline{\ln p} \, . \tag{3.53}$$

The final form on the RHS is appropriate, since the mean probability $p(\alpha)$ for individual microscopic states and for given $\rho(\alpha)$ is then $p(\alpha) = \rho(\alpha)/\sigma(\alpha)$. The entropy defined by $-kH_{\text{gen}}$ is also known as the *relative entropy* of $\rho(\alpha)$ with respect to the measure $\sigma(\alpha)$. The latter is often introduced *ad hoc* as part of a phenomenological description.

Under the approximation $w(\alpha', \alpha) = f(\alpha)\delta'(\alpha - \alpha')$ one now obtains the deterministic 'drift' limit of the master equation (3.51) – usually representing the first term of (3.44). It defines the first order of the *Kramers–Moyal expansion* for $w(\alpha, \alpha')$, equivalent to an expansion of $\rho(\alpha', t)$ in terms of powers of $\alpha' - \alpha$ at $\alpha' = \alpha$. The second order, $w(\alpha', \alpha) = f(\alpha)\delta'(\alpha - \alpha') + g(\alpha)\delta''(\alpha - \alpha')$, leads to the *Fokker–Planck equation* as the lowest non-trivial approximation that leads to an irreversible equation (see de Groot and Mazur 1962, Röpke 1987). In this respect, it is analogous to the LAD equation as the lowest non-trivial order in the Taylor expansion of the Caldirola equation (2.31). A master equation is generally equivalent to a (stochastic) *Langevin equation* for *individual* macroscopic trajectories $\alpha(t)$ which may form a dynamical ensemble represented by $\rho(\alpha, t)$.

In contrast to the Liouville equation (3.26), the master equation (3.48) or (3.36) cannot be unitary with respect to the inner product for probability distributions defined above (3.27). While total probability must be conserved by these equations, that of the individual trajectories *cannot* (see also Sect. 3.4). Information-reducing master equations describe an indeterministic evolution, which in general only determines an ever-increasing ensemble of different *potential* successors for each macroscopic state (such as a point in α-space).[7] As discussed above, this macroscopic indeterminism is compatible with microscopic determinism if that information which is transformed from relevant

---

[7] The frequently used picture of a 'fork' in configuration space, characterizing a dynamical indeterminism, may be misleading, since it seems to imply unique predecessors. This would be wrong, as can be recognized, for example, in an equilibrium situation. In the case of a stochastic dynamical law that is defined on a finite set of states, a state must in general also have different possible predecessors, corresponding to an inverse fork. Inverse forks by themselves would represent a pure forward determinism (a 'semigroup', that may describe *attractors*). All these structures are meant to characterize the *dynamical law*. They are neither properties of the (f)actual history (which is assumed to evolve along a *definite* trajectory regardless of the nature of the dynamical law), nor of an evolving ensemble that represents a specific state of knowledge.

However, only dynamically unique predecessors may give rise to recordable histories (consisting of 'facts' that are redundantly documented). The *historical nature* of our world is thus based on a uniquely determined or even overdetermined macroscopic past – see also footnote 1 of Chap. 2, Fig. 3.8, and Sect. 3.5. A macroscopic history that was completely determined from its macroscopic *past* would be in conflict with the notion of an (apparent) free will.

to irrelevant in the course of time no longer has any relevant (macroscopic) effects for all future times of interest. The validity of this assumption depends on the dynamics and on the specific initial conditions (3.46).

Time-reversed ('anti-causal') effects could only be derived from an appropriate *final* condition by applying the corresponding approximations to (3.44) for $t < t_0$. It is an *empirical fact* that such a condition, analogous to $A^{\mu}_{\text{out}} = 0$ in electrodynamics, does not describe our observed Universe. An exact boundary condition $\rho_{\text{irrel}}(t_0) = 0$ at some *accessible time* t_0 would for similar statistical reasons lead to a non-*de*creasing entropy for $t > t_0$, but to non-*in*creasing entropy for $t < t_0$, hence to an entropy minimum at $t = t_0$ unless $S(t_0) = S_{\max}$.

While the (statistically probable) assumption (3.46) led to the master equation (3.48), it would *not* necessarily characterize an arrow of time. Without an improbable initial condition $\rho_{\text{rel}}(t_i)$, the approximate validity of the equality sign in (3.50) would be overwhelmingly probable. Retarded action-at-a-distance electrodynamics would be trivial, too (and equivalent to its advanced counterpart) if *all* sources were already in thermal motion (such as the sources forming absorbers). It is the low entropy initial condition for ρ_{rel} which is responsible for the dynamical formation of that 'irrelevant' information which would be highly relevant for correctly calculating $\rho_{\text{rel}}(t)$ *backwards* in time.

The main conclusions derived in this and the previous section can thus be summed up as follows:

1. The *ensemble entropy* S_Γ does not represent physical entropy, since (a) it would be minus infinity for a real physical state (one or $N!$ points in phase space), (b) it is otherwise not additive for composite systems (in particular, it is not an integral over an entropy density), and (c) it remains constant under deterministic dynamics (in contrast to the Second Law). For indeterministic dynamical laws, it would have to increase, starting from its given value, in *both* directions of time (except when already at its maximum value). This demonstrates that ensemble entropy is *not* a physical quantity (see also Kac 1959).
2. Coarse-grained (or 'relevant') entropy, when defined as a function of the deterministically evolving microscopic state that is assumed to represent reality, would *most probably* fluctuate in time close to its maximum value. However, it may increase for a very long time – far exceeding the present age of the Universe if this had begun in an appropriate state of extremely low entropy (see Sect. 5.3). While a Zwanzig projection (describing generalized coarse-graining) can be arbitrarily *chosen for convenience* in order to derive an appropriate master equation, the cosmic initial condition must be *specified* as a condition characterizing the real Universe.
3. Only a relevance concept that includes locality is able to describe entropy as an extensive quantity.

4. Any coarse-grained entropy could be *forced* never to decrease by an appropriate modification of the corresponding *ensemble* dynamics – as in (3.36). This may represent either new physics or an approximation to the situation described in the second part of item 2 (where the second possibility is assumed to apply).

General Literature: Jancel 1963, Balian 1991.

3.3 Thermodynamics and Information

3.3.1 Thermodynamics Based on Information

As explained in the previous sections, Gibbs' probability densities or ensembles ρ_Γ represent incomplete *information* about the real state, which would in classical mechanics be described by a singular point in phase space. Similarly, Zwanzig's projection operators $\hat{P}$ (defining a generalized coarse-graining) were justified by the incomplete observability of macroscopic systems. The entropy and other parameters characterizing these ensembles, such as a temperature, therefore appear fundamentally observer-related (objectively unmotivated). While Gibbs' ensembles refer in principle to *actual* knowledge, Boltzmann's distributions may be based on an objectivized limitation of knowledge, characterizing a certain class of potential observers, such as those able to recognize only the mean particle density ρ_μ for a gas.[8] For similar reasons, the coarse-graining $\hat{P}$ is kept fixed as a reference system, and not comoving according to the dynamics. The concept of information appears here extraphysical, although observers or other carriers of information have to be regarded as physical (in particular thermodynamical) systems, too (see Sect. 3.3.2).

Jaynes (1957) generalized Gibbs' statistical methods by rigorously applying Shannon's (1948) information concept. Shannon's formal measure of information for a probability distribution $\{p_i\}$ on a set of elements characterized by the index i,

$$I := \sum_i p_i \ln p_i \leq 0 \; , \tag{3.54}$$

is evidently defined in analogy to Boltzmann's H, and therefore also called *negentropy*. However, as a measure of information, it corresponds more closely to Gibbs' extension in phase η. This measure is often normalized *relative* to its value for minimum information, $p_i = p_i^{(0)}$, where $p_i^{(0)} = 1/N$ if $i = 1, \ldots, N$, unless different statistical weights for the 'elements' i arise from a more fundamental level of description – cf. (3.53):

[8] The term 'objectivized' presumes the basically subjective (observer-related) status of what is to be objectivized. In contrast, the term 'objective' is in physics often used synonymously with the term 'real', and then means the assumed or conceivable existence of an object or its state regardless of its observation.

$$I_{\rm rel} = I(p_i|p_i^{(0)}) := \sum_i p_i \ln(p_i/p_i^{(0)}) = \ln N + \sum_i p_i \ln p_i \geq 0 \ . \tag{3.55}$$

This renormalized measure of information may remain finite even when I diverges in the limit $N \to \infty$. Under an appropriate modification it can then also be applied to a continuum.

Jaynes thus based his approach on the idea that the microscopic state of a macroscopic system can never be completely *known*. Instead, a small though varying number of macroscopic variables, which are functions of the microscopic state, $\alpha(p,q)$, are approximately 'given'. Therefore, he introduced specific representative ensembles, $\rho_\alpha(p,q)$ or $\rho_{\bar{\alpha}}(p,q)$, which are defined to possess minimal information about all other variables (maximal ensemble entropy $S_\Gamma[\rho]$) under the constraint of either fixed values α, or fixed *mean* values $\bar{\alpha} := \int \alpha(p,q)\rho(p,q)\mathrm{d}p\,\mathrm{d}q$. This entropy thus becomes a *function* of α or $\bar{\alpha}$, defined as $S(\bar{\alpha}) := S_\Gamma[\rho_{\bar{\alpha}}]$, for example. This generalization of Gibbs' approach has turned out to be useful in many applications, while the macroscopic variables α remain to be chosen *ad hoc*.

As mentioned already in Sect. 3.1.2, an entropy concept based on the actually available information would be in conflict with the usual interpretation of entropy as an observer-independent physical quantity that can be objectively measured. On the other hand, its dependence on a certain basis of information may be quite meaningful. For example, the numerical value of $S_\Gamma[\rho]$ depends in a reasonable way on whether or not ρ contains information about actual density fluctuations, or about the isotopic composition of a gas. The probability $p_{\rm fluct}(\alpha)$ for the occurrence of some quantity α in thermodynamical equilibrium was successfully calculated by Einstein in his theory of Brownian motion from the expression

$$p_{\rm fluct}(\alpha) = \frac{\exp\left[S(\alpha)/k\right]}{\exp\left\{S[\rho_{\rm can}]/k\right\}} \ , \tag{3.56}$$

thus exploiting the interpretation of entropy as a measure of probability. The probability for other quantities to be found immediately after the observation of this fluctuation would then have to be calculated from the 'conditioned' ensemble ρ_α rather than from $\rho_{\rm can}$.

Similarly, a star cluster (that is, a collection of macroscopic objects) possesses meaningful temperature and entropy $S \neq 0$ from the *point of view* that the motion of the individual stars is regarded as 'microscopic'. The same statistical considerations as used for molecules then show that their velocity distribution must be Maxwellian. At the other extreme, one could (in classical physics) conceive of an external Lapacean demon as a *super-observer* of the individual molecules in a gas. Entropy would indeed depend here on the available or accessible information. Its objectivity in thermodynamics can then only be understood as representing a common basis of information shared by us human observers. This perspective must be caused by our specific situation as physical systems.

In order to consistently regard ensembles as representing *actual* information, one would have to take into account all physical processes which affect the information carrier rather than just those in the system itself. Such a definition would certainly be inappropriate for the concept of physical entropy. For example, thermodynamical entropy does not depend on whether or how accurately the temperature has been measured; it is simply understood as a function of temperature.

Let $\alpha(p,q)$ represent a set of such quantities that are assumed to be 'given', possibly up to certain uncertainties $\Delta\alpha$ – see the model used in (3.51). The Hamiltonian $H(p,q)$ is in general just one of them. Subsets of microscopic states p,q corresponding to values of α within intervals $\alpha_0 < \alpha(p,q) < \alpha_0 + \Delta\alpha$, define subvolumes of $\mathit{\Gamma}$-space. The widths $\Delta\alpha$ may be those of Jaynes' representative ensembles for given mean values $\bar{\alpha}$, since any finer resolution would regard fluctuations as being relevant, unless α were a constant of the motion. For a single parameter α, these volume elements can be written as $\Delta V_\alpha := (\mathrm{d}V/\mathrm{d}\alpha)\Delta\alpha$, with $V(\alpha_0) := \int_{\alpha(p,q)<\alpha_0} \mathrm{d}p\,\mathrm{d}q$. In N-particle phase space, the size of the interval $\Delta\alpha$ is often quite irrelevant, since contributions to the volume integral for a compact region $\alpha(p,q) < \alpha_0$ may be strongly concentrated just below the surface defined by the value α_0 because of the geometry of such high-dimensional spaces. The term $\ln \mathit{\Delta}\alpha$ can then be neglected under the logarithm, $\ln \mathit{\Delta} V_\alpha$, that defines the entropy $S(\alpha)$.

One may now define a new useful Zwanzig projection $\hat{P}_{\mathrm{macro}}$ by *averaging* over subsets defined by such volume elements ΔV_α:

$$\hat{P}_{\mathrm{macro}}\rho(p,q) := \frac{\Delta p_\alpha}{\Delta V_\alpha} \qquad (3.57)$$
$$:= \frac{1}{\Delta V_\alpha}\int_{\Delta V_\alpha} \rho(p',q')\mathrm{d}p'\mathrm{d}q' , \quad \text{for} \quad p,q \in \Delta V_\alpha .$$

If discrete values α_i are defined for convenience by means of 'macroscopic steps' $\alpha_i + \Delta\alpha = \alpha_{i+1}$, the integral for $S_\Gamma[\hat{P}_{\mathrm{macro}}\rho]$ splits into two sums:

$$S_\Gamma[\hat{P}_{\mathrm{macro}}\rho] = -k\int \hat{P}_{\mathrm{macro}}\rho \ln(\hat{P}_{\mathrm{macro}}\rho)\mathrm{d}p\,\mathrm{d}q$$
$$= -k\sum_i \Delta V_{\alpha_i}\frac{\Delta p_{\alpha_i}}{\Delta V_{\alpha_i}} \ln\frac{\Delta p_{\alpha_i}}{\Delta V_{\alpha_i}}$$
$$= -k\sum_i \Delta p_{\alpha_i}\ln \Delta p_{\alpha_i} + \sum_i \Delta p_{\alpha_i} k \ln \Delta V_{\alpha_i} . \qquad (3.58)$$

[Note the relation to the concept of relative information (3.53) or (3.55) – see also Schlögl 1966.] The first term in the last line describes the entropy corresponding to the lacking *macroscopic information* described by the probabilities Δp_{α_i}. The second term is the *mean physical entropy* with respect to this macroscopic ensemble. The physical entropy, $S(\alpha) := k\ln \Delta V_\alpha \approx S_\Gamma[\rho_\alpha]$, thus measures the size of Jaynes' representative ensembles ρ_α, or, in Planck's language, the *number of complexions*, that is, the number of microscopic states which may represent it. In the special case $\alpha(p,q) := H(p,q)$ one obtains the entropy of the canonical ensemble as a function of the mean energy. If $\Delta\alpha = \Delta E$ is chosen infinitesimal, one obtains the entropy of the microcanonical ensemble, relevant for thermodynamically closed systems.[9]

Although the first term on the RHS of (3.58) is usually much smaller than the second one, it is essential for a complete and consistent discussion of information processing and measurement (see Sect. 3.3.2). A simple example of such a partitioning of the ensemble entropy into physical entropy and entropy of lacking information is provided by the particle number in a *grand* canonical ensemble, $Z^{-1} \exp\left[-(H - \mu N)/kT\right]$. This particle number is assumed to be 'given' (although in general not known) once the vessel that was in equilibrium with a particle reservoir characterized by the chemical potential μ has been closed. Thereafter, the system is represented by a *canonical* ensemble with fixed particle number N, while the relative contribution of that part of the original ensemble entropy which has now become entropy of lacking information about the exact particle number N is of the order $\ln N/N$ (Casper and Freier 1973). This contribution to the entropy is often neglected by using the 'approximation' $N! \approx N^N$. The argument demonstrates, however, that this different choice of ensembles is dynamically justified (by their robustness), and that the difference between the number of permutations, $N!$, of a fixed number N of particles and the factor N^N arising from the grand canonical ensemble with *mean* particle number N is meaningful – see (4.21) and cf. footnote 2.

The concept of physical entropy, defined above, no longer depends on actual information, since the choice of 'macroscopic' subsets, characterized by functions of state $\alpha(p,q)$, is motivated by their dynamical stability. In general, variables α characterizing 'robust' subsets of phase space that are densely populated by a trajectory (in the sense of quasi-ergodicity) within reasonably short times are regarded as macroscopic quantities. This quasi-ergodicity depends on a 'measure of distance' in Γ-space that cannot be invariant under canonical transformations. The macroscopic variables α are instead assumed to vary slowly and controllably – even under the influence of normal perturbations, or during their observation. These robust quantities define approximate constants of the motion or adiabatically changing collective variables. Since this concept of robustness is based on quantitative aspects, it cannot usually be defined with mathematical rigor. For example, the positions and shapes of droplets that are formed in a condensation process, or even more so those of the walls of the vessel, are evidently robust properties, although they do not represent exact constants of the motion.

9 The infinite renormalization which is required for the corresponding concept of an entropy *density* as a function of α is due to the fact that the entropy for a continuous quantity has no lower bound, so that the measure of information may grow beyond all limits – see the remarks following (3.55).

A microscopic trajectory $q(t)$ determines all macroscopic trajectories $\alpha(t)$ defined as functions of this state: $\alpha(t) := \alpha\big(p(t), q(t)\big)$. As discussed in Sect. 3.1.2, the macroscopic dynamics $\alpha(t)$ is then in general not autonomous, since trajectories starting from the same $\alpha(t_0)$ may evolve into different $\alpha(t_1)$ – depending on the microscopic initial state $p(t_0), q(t_0)$. This *macroscopic indeterminism* is essential for fluctuations or certain phase transitions.

The determinism of a dynamical *model* (such as Laplacean mechanics) is defined by the *mathematical existence* of a unique mapping of appropriate initial (or final) states onto complete trajectories. This concept of determinism is independent of the availability of an (analytic or algorithmic) *procedure* for explicitly constructing these trajectories in terms of conventional coordinates (‘integrability’). It is therefore also independent of any *practical* limitation to their computability, which forms the basis of Kolmogorov’s (1954) entropy, and is often used in the definition of *chaos* (see Schuster 1984, or Hao-Bai-Lin 1987). In classical mechanics, the deterministic dynamical mapping of initial conditions onto trajectories is a consequence of Newton’s equations under non-singular conditions (see Bricmont 1996 for his lucid criticism of the popular misuse of the concept of chaos in this connection).

Trajectories could in principle be described in terms of the constants of the motion. The latter could then be used as new coordinates or ‘co-evolving grids’ (see Appendix B of Zurek 1989). Such constants of the motion are often denied to exist, since they are not analytically related to conventional coordinates. However, this does *not* mean that they would not exist in any absolute sense. It was indeed one of the great lessons from the theory of relativity that physics and spacetime geometry (‘reality’) are independent of the choice of coordinates, while the ancient Greeks were not even able to overcome Zeno’s paradox of Achilles and the tortoise by a transformation to more appropriate ‘coordinates’ of description. We should similarly be able to conceptually overcome all mathematical problems in the construction of canonical transformations, and instead rely on the assumption of a coordinate-free ‘reality’ (at least in classical mechanics).

These mathematical difficulties may nonetheless reflect the complex and non-trivial *physical* relation between the Universe and its ‘observing parts’. Observers are evidently *not* in any simple way related to the constants of the motion – the reason why we feel ‘time change’.[10] Some authors have related the problems of a universe that contains its observers (*physical* self-reference) to Gödel’s undecidability theorems, which apply to logical systems that allow *formal* self-reference (see Wheeler 1979). However, one cannot argue that the existence or meaning of an observer-independent reality is excluded just because of the observers’ limited capabilities. This insufficient argument has even been used as an explanation of ‘quantum uncertainty’ (Popper 1950, Born 1955, Brillouin 1962, Cassirer 1977, Prigogine 1980). There is a fundamental

[10] “Time goes, you say? Ah no! Alas, time stays, we go.” (Austin Dobson – discovered in Gardner 1967.)

difference between the *impossibility of ever knowing* the precise classical state of the Universe and the *incompatibility of its existence with certain empirical facts*. While the former is often derived precisely by using (thus presuming) classical concepts as describing reality, the latter is a consequence of crucial experiments on which quantum theory is based.

3.3.2 Information Based on Thermodynamics

Macroscopic indeterminism, such as described by Einstein's fluctuations in (3.56), may give rise to a transient *decrease* in physical entropy $S(\alpha)$ in accordance with microscopic determinism. It requires the transformation of lacking irrelevant into lacking relevant information. The latter would *not* be lacking any more if the fluctuation were observed, or, similarly, after the *measurement* of a microscopic variable, as depicted by the first step of Fig. 3.5. In these cases, the physical realization of information by observers or other information carriers has to be properly taken into account.

As is well known since the discussion of *Maxwell's demon*, any change or use of information must be described physically, with all its thermodynamical consequences. Maxwell had assumed his demon to operate a microscopic sliding door between two compartments of a vessel in such a way that only fast molecules may enter the first compartment, while only slow ones are allowed to leave it. His actions must then lead to a temperature and pressure difference, thus admitting the construction of a perpetuum mobile of the second kind.

The demon must here invest *its knowledge* about trajectories of individual molecules. However, Smoluchowski (1912) objected that a demon who acts physically would itself have to obey the Second Law: its operations must be described (thermo-)dynamically. In phenomenological terms, any lowering of the entropy of the gas must at least be balanced by a corresponding increase of the demon's entropy. If the demon were assumed to be a finite and thermodynamically closed system, its increasing Brownian motion would then ultimately prevent it from acting properly (by letting its 'hands tremble', or as a result of its deteriorating information about the molecules).

Szilard (1929) derived a fundamental information-theoretical consequence from this situation. By exploiting the idea of Maxwell's demon, he concluded that an 'intelligent being' must use up an amount of information of measure

$$\Delta I = \frac{\Delta S}{k} \,, \tag{3.59}$$

in order to *lower* the entropy of some system by ΔS. This equivalence would also be compatible with the ensemble interpretation of entropy, or Einstein's probabilities (3.56).

Szilard's main argument used a model 'gas' consisting of a single molecule in a vessel of volume V. Statistical aspects are introduced by means of many collisions of the molecule with the walls, leading to thermal equilibration

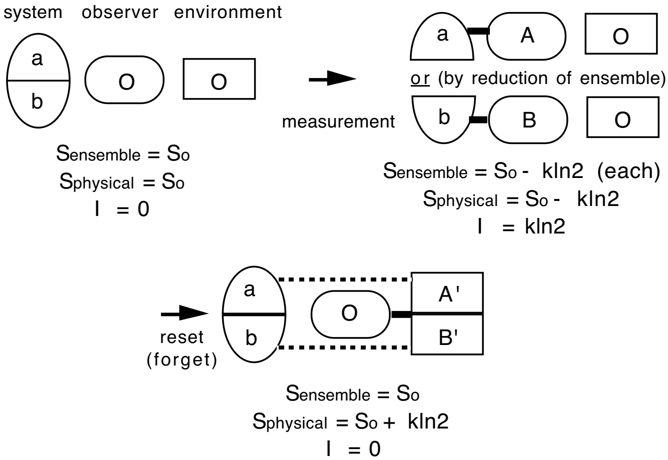

Fig. 3.5. Entropy relative to the state of information during a classical measurement. In the first step in the figure, the state of the observer changes depending on that of the system. The second step represents the subsequent resetting of the 'observer' or device (Bennett 1973), required if the process is to be exactly repeated for a second measurement. Areas represent sets of microscopic states of the subsystems (while those of uncorrelated combined systems would be represented by their direct products). The lower case letters a and b characterize the property to be measured; and 0, A and B the corresponding 'memory states' of the observer, while A' and B' are their respective effects in the thermal environment, required for a deterministic reset. The 'physical entropy' (*defined* to add for subsystems) measures the phase space of all microscopic degrees of freedom, including the property to be measured. Because of this presumed additivity, the physical entropy neglects statistical correlations (*dashed lines*, which indicate *sums* of direct products of sets) as being 'irrelevant' in the future – hence $S_{\text{physical}} \geq S_{\text{ensemble}}$. I is the amount of information held by the observer. S_0 is at least $k \ln 2$ in this simple case of two equally probable values a and b. (From Chap. 2 of Joos et al. 2003)

between the molecule's average motion and a surrounding heat bath (see Fig. 3.6). A piston is then inserted sideways (without using energy) in order to separate two partial volumes V_1 and V_2. This partition of the volume is *robust* in the sense of Sect. 3.3.1. According to (3.58), this procedure transforms part of the (physical) entropy of the 'gas' into entropy of *lacking information*. If the experimenter knows (only) in which partial volume $i = 1, 2$ the molecule resides, corresponding to a Shannon measure $\Delta I_i = \ln[(V_1 + V_2)/V_i]$, he is able to retrieve the mechanical energy

$$\Delta A_i = \int_{V_i}^{V_1+V_2} p \mathrm{d}V = \int_{V_i}^{V_1+V_2} \frac{kT}{V} \mathrm{d}V = -kT \ln \frac{V_i}{V_1 + V_2} \tag{3.60}$$

by moving the piston into the empty volume, and slowly raising a weight, for example. The molecule's mean kinetic energy may thereby remain constant by

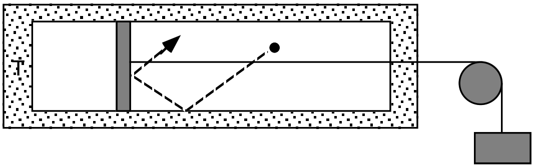

Fig. 3.6. Szilard's *Gedanken* engine completely transforms thermal energy into mechanical energy by using information

the reversible extraction of heat from the external reservoir with temperature T. This process lowers the entropy of the reservoir by

$$\Delta S_i = -\frac{\Delta A_i}{T} = k \ln \frac{V_i}{V_1 + V_2} = -k\Delta I_i \ , \tag{3.61}$$

in accordance with (3.59).

According to Smoluchowski, one could avoid referring to knowledge or information by using a 'mechanical rectifier' (such as a ratchet) that causes the piston to move in the appropriate direction. This rectifier would ultimately have to perform thermal motion large enough to make it useless, corresponding to the demon's trembling hands (see also Feynman, Leighton and Sands 1963, Vol. I, p. 46-1). So one has to conclude that utilizing knowledge for making decisions (for example in the brain) is equivalent to the operation of a rectifier. It is here essential that the rectifier cannot be reset to its initial state without getting rid of entropy – usually in the form of heat (Bennett 1987). For this reason the mechanism cannot work reversibly in a closed system.

Brillouin (1962), when elaborating on ideas originally presented by Gabor in lectures given in 1952 (see Gabor 1964), emphasized that Szilard's 'intelligent being' has to *acquire* information. Since this process must also be compatible with the Second Law, Brillouin postulated his *negentropy principle*

$$\Delta S' - k\Delta I \geq 0 \ , \tag{3.62}$$

which meant that any information gain ΔI has to be accompanied by some process of dissipation that leads to a *production* of thermodynamical entropy $\Delta S'$ *in the information medium* (usually light). He thereby referred to the latter's quantum aspect (photons), which limits its information capacity. Because of the minimum information required according to Szilard, the construction of a perpetuum mobile of the second kind would then again be excluded. However, because of the above example of a directly coupled mechanical rectifier, no explicit reference to an information medium seems to be required. According to Bennett (1973), it is the increase in physical entropy by the reset mechanism in Fig. 3.5, $\Delta S_{\text{phys}} = k \ln 2$ if $V_1 = V_2$, that compensates its decrease in (3.61).

All non-phenomenological arguments are based here on two assumptions: (1) Global determinism, which requires that an ensemble of N different states (or N ensembles of equal measure) must have N different successors, which have to be counted by the total ensemble entropy. Different states may evolve into the *same* final state only by means of an appropriate interaction with their environment, that transfers this difference to the latter (for example, in the form of heat). (2) Intuitive causality, which asserts that uncontrollable 'perturbations' by the environment can only *enlarge* the ensemble. It gives rise to *in*equalities such as (3.62) rather than equations. If thermodynamical concepts apply, a transfer of entropy ΔS must be accompanied by a transfer of energy according to $\Delta Q = T\Delta S$. This relation has also led to the interpretation of entropy as a measure of *degradation of energy*.

The equivalence of information and negative entropy suggests that any (tautological) information *processing* (for example in a computer) can in principle be performed reversibly. However, arithmetic operations are often *logically irreversible* in the sense that two factors cannot be recovered from their product. (In a mechanical computer this operation may indeed require friction.) This led to the conjecture that a minimum amount of entropy $k \ln 2$ has to be produced for each bit of information in each elementary calculational step (Landauer 1961). It was refuted by Bennett (1973 – see also Bennett and Landauer 1985). However, in their discussion the logically lost information ('garbage bits') – even if randomized – is still regarded as macroscopic or 'relevant' in the thermodynamical sense. For this reason, the entropy creation is deferred to the reset or clearing of the memory, which is required for the computer to perform its calculational steps more than once (see the second step of Fig. 3.5). These considerations will lead to quite novel consequences for quantum computers (see Sect. 4.3.3).

All these arguments support the interpretation that information has to be *physically* realized (and therefore to be compatible with the laws of thermodynamics), rather than representing an extraphysical concept that has to be independently postulated for a statistical foundation of thermodynamics. On the other hand, *mathematical theorems* do *not* represent information (as, for example, assumed by Landauer 1996). Logic deals exclusively with tautologies ('analytical judgements') – as complicated as they may appear to our limited intelligence.

General Literature: Denbigh and Denbigh (1985), Bennett (1987), Leff and Rex (1990).

3.4 Semigroups and the Emergence of Order

In physical systems, ‘ordered’ states are characterized by low entropy. Order may appear in the form of simple structures (such as regular lattices) or complex ones (organisms). For example, the rectifier discussed in the previous section as replacing Maxwell’s demon must display ordered *dynamical* behavior. The emergence of order from disorder in Nature, also called *self-organization of matter*, may appear to contradict the Second Law with its general trend towards disorder and *chaos*. This has often been misunderstood as a ‘discrepancy between Clausius and Darwin’. However, the fundamental phenomenological equation (3.1) allows entropy to decrease *locally*. A negative first term would allow physical entropy to flow into the environment. If this environment is not in complete thermal equilibrium, and characterized by at least two different temperatures, T_1 and T_2, a local loss of entropy, $\mathrm{d}S_{\mathrm{ext}} = \mathrm{d}Q_1/T_1 + \mathrm{d}Q_2/T_2 < 0$, would not even require any net loss of heat, $\mathrm{d}Q_1 + \mathrm{d}Q_2 < 0$. (Here differentials are always meant to refer to positive time increments $\mathrm{d}t$.) This local decrease of entropy is thus *not* in conflict with its global increase according to the Second Law – see also Sect. 5.3.

In statistical terms, the number of states in a *dynamically representative ensemble* (see Sect. 3.1.2) may decrease locally in accordance with determinism *and* intuitive causality, provided the ensemble characterizing the state of the environment increases accordingly – precisely as during the ‘reset’ of a memory device, indicated in Fig. 3.5. In this Laplacean description, the outcome of evolution would be determined by the microscopic initial state of the whole Universe.

An important special case is a *steady state* of non-equilibrium, characterized by $\mathrm{d}S = \mathrm{d}S_{\mathrm{int}} + \mathrm{d}S_{\mathrm{ext}} = 0$ in spite of non-vanishing entropy production, $\mathrm{d}S_{\mathrm{int}} > 0$ (Bertalanffi 1953). It may support ordered states as *dissipative structures*. The standard example, known as *Bénard’s instability*, describes convective heat transfer through a thin horizontal layer of a liquid in the form of spatially ordered convection cells, which optimize the process of thermal equilibration between two reservoirs at different temperatures. In a finite universe, this stationary situation can only represent a transient local phenomenon. The emergence of structure is often connected with symmetry breaking (in particular of translational symmetry), related to a phase transition. In a deterministic description, an initial microscopic fluctuation would thereby become unstable and be amplified to a macroscopic scale. In quantum theory, it may also require an indeterministic collapse of the wave function (see Sect. 4.1.2).

For similar reasons, Boltzmann suggested that biological processes here on earth are facilitated by the temperature difference between the sun (with its 6000 K surface temperature) and the dark Universe (at 2.7 K, as we know today). At the distance of the earth, the solar radiation has an energy density much lower than that of a black body with the same spectrum (temperature). Since photon number is not conserved (in general not even a robust quantity), a canonical distribution $\exp(-H/kT)$ in the occupation number representation determines not only the spectral distribution as a function of temperature, but also the intensity (photon density). A gas with conserved particle number would instead allow one independently to choose the mean density – either by fixing the particle number by closing the vessel, or by fixing the chemical potential (in a grand canonical ensemble) by connecting the vessel to a particle reservoir. In contrast, a photon from the sun can be transformed very efficiently into many soft photons, which together possess much higher physical entropy.

Although order appears to be an objective property, an absolute concept of order that is not simply defined by means of phenomenological entropy is as elusive as an objective concept of information or relevance (see Denbigh 1981, p. 147, or Ford 1989). For reasons already mentioned in Sect. 3.3.1, the definition of order in terms of computability would depend on the choice of 'relevant coordinates'. For example, the obvious order observed in a crystal lattice is not invariant under general canonical transformations. How, then, may the order of an organism be conceptually distinguished from the 'chaotic' correlations arising from molecular collisions in a gas?

Many self-organizing systems include chemical reactions. They are phenomenologically described by irreversible rate equations, which define the dynamics of concentrations $X, Y, \ldots$ These concentrations are 'macroscopic' variables, called α in Sects. 3.2 and 3.3.1. In statistical terms, rate equations can be derived from a generalized *Stoßzahlansatz* that includes rearrangement collisions between different kinds of molecules, which are usually assumed to be already in *thermal* equilibrium with one another. These rate equations are therefore special master equations (as derived in Sect. 3.2) for these 'relevant' degrees of freedom $X, Y, \ldots$

Rate equations determine trajectories in the configuration space of concentrations.[11] For closed systems, these trajectories may eventually approach that point in their configuration space which describes equilibrium. Reversible determinism must come to an end at such *attractors* (see Fig. 3.7a), although this may require infinite time. A *mechanical* example of an attractor in the presence of friction is the phase space point characterized by $v := \mathrm{d}x/\mathrm{d}t = 0$ and $V(x) = V_{\min}$. The corresponding equation of motion, $m\mathrm{d}v/\mathrm{d}t = -av - \nabla V$, neglects any stochastic response from the energy-absorbing microscopic degrees of freedom, which is in principle required by the *fluctuation–dissipation theorem*. Similar to the LAD equation of Sect. 2.3, this equation is, therefore, deterministic, even though it is asymmetric under time reversal.

Points in the space of macroscopic variables X, Y or x, v ('macroscopic states' α, in general) describe the physical states incompletely. They represent large subspaces of the complete Γ-space (that may realistically even have to include the environment). Volume elements of the same size in macroscopic

[11] As the rate equations are of first order in time, this macroscopic configuration space is often called a phase space.

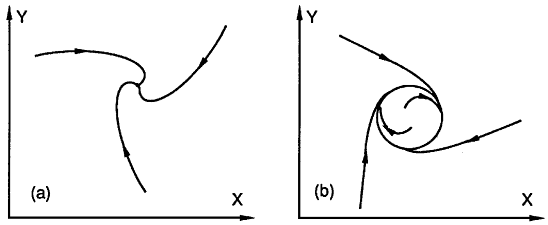

Fig. 3.7. Standard representation of an attractor (**a**) and a limit cycle (**b**) as examples of phenomenologically irreversible dynamics in the configuration space of macroscopic variables $\alpha \equiv X, Y$

'phase space' may correspond to very different ensemble measures. These volume elements are therefore in general not dynamically conserved. For example, the immediate vicinity of an equilibrium 'state' X_0, Y_0 – such as $v = 0$, $V(x) = V_{\min}$ in the mechanical example – covers almost the whole Γ-space of the completely described system (or some subspace that is defined by conserved quantities).

In the specific mechanical example with friction, the *modified* macroscopic phase space measure $\mathrm{d}x\,\mathrm{d}v/v$ nonetheless happens to be dynamically invariant. Time reversal is here compensated for by a transformation $v \to 1/v$ to restore a formal T-symmetry of macroscopic determinism (in formal analogy with the examples mentioned in the Introduction). This leads to a conserved *generalized H-functional* – cf. (3.53) and (3.55), viz.,

$$H_{\text{gen}} := \int \rho(v, x) \ln \left[|v|\, \rho(v, x) \right] \mathrm{d}v\, \mathrm{d}x \,, \tag{3.63}$$

which defines a 'reference density' $\rho_0 = |v|^{-1}$ as an effective equilibrium measure on this macroscopic phase space.

In the situation of a steady state non-equilibrium, macroscopic trajectories described by other effective irreversible equations of motion may approach certain *closed curves*, which do *not* correspond to maximum entropy. They are called *limit cycles*, and may represent dissipative structures, which represent order (see Fig. 3.7b and Glansdorff and Prigogine 1971).

Open systems are often described by means of *phenomenological semigroups*, defined as dynamical maps acting on ensembles in finite time steps. These maps can be understood as time-integrated Markov operators G_{ret}, and are thus applicable again only in the 'forward' direction of time (in contrast to the reversible group of time translations, valid for dynamically closed systems). Mathematically, ensembles may even be regarded as the fundamental kinematical objects of the theory, without any explicit definition of their *elements*, which would describe microscopic reality. 'Determinism' is then understood as a mere forward determinism for these formal ensembles. *Maps* are called irreversible if they form genuine semigroups, that is, if they cannot be uniquely inverted *as maps on ensembles*.

This irreversibility *of maps* does *not* correspond to a dynamical indeterminism for elementary states: it usually represents resets or attractors phenomenologically – that is, without explicitly taking into account microscopic degrees of freedom. In order to describe a reset, the master equation (3.48) has to be based on a *non-Hermitean* Zwanzig projector that 'creates' relevant information. In a globally deterministic context, its microscopic realization would then have to contain some way of getting rid of entropy (as discussed in Sect. 3.3.2).[12] As can be seen from the second step of Fig. 3.5, the reset transforms local information into nonlocal correlations (also depicted in Fig. 3.1). This transformation describes a *production of physical entropy*, while the ensemble entropy is conserved. The absence in Nature of correlations which would allow the inverse process and thus lead to a reduction of physical entropy is responsible for the irreversibility of the semigroup.

As these semigroups are defined to act on ensembles, regarded as abstract objects, their inversion does *not* in general represent a reversal of the microscopic dynamics ('time reversal'). For the same reason, their forward determinism is not equivalent to microscopic determinism. A dynamical map may not be invertible *as a map* even though the underlying dynamical transformation of microscopic states can be reversed.

Individual indeterminism and attractors are illustrated on a finite set of states in Fig. 3.8. An *asymmetric* dynamical indeterminism (b) is represented by diverging forks (see footnote 7), while an attractor is characterized by converging (or 'inverse') forks (c). An everywhere defined indeterminism must apply symmetrically (a). On a continuum of states, one would first have to define a measure, usually according to its invariance under the assumed *fundamental* deterministic dynamics of the completely described closed system. (This may represent a problem if determinism is to be given up fundamentally.) Semigroups are often studied on discrete state spaces, where measures of states are trivial. A popular example is the model of 'deterministic cellular automata' (see Kauffman 1991). Their merging trajectories (representing attractors) then replace the shrinking phase space in the continuum model with friction that led to the generalized H-functional (3.63).

Forward-deterministic dynamical maps are often defined by means of nonlinear transformations. A popular (though not very physical) one-dimensional toy model of a semigroup is the *Bernoulli shift*, defined by the mapping

[12] Therefore, mathematical physicists have proposed a new definition of entropy that would *always* allow the entropy of an open system to grow under a semigroup, even if its physical entropy decreased – see Mackey (1989). For example, the *relative entropy* – cf. (3.53) and (3.55)], defined for an open system with respect to a canonical distribution with temperature of an *external heat bath*, would increase even when the temperature of the system is *lowered* (as it is in the case of a cooler heat bath). Such a formal redefinition of entropy is certainly physically misleading, even though it may be useful for certain purposes.

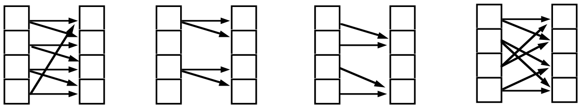

Fig. 3.8. Dynamical transformations of states on a discrete and finite 'phase space' consisting of only four states: (**a**) T-symmetric indeterminism (representing an incompletely determined Hamiltonian, for example); (**b**) asymmetric indeterminism, representing a law-like increase of ensemble entropy (cannot be defined everywhere on finite sets); (**c**) attractors (cannot be inverted as a map defined on all ensembles); (**d**) discrete caricature of a Frobenius–Perron map (see text). The symmetric indeterminism (**a**) would appear asymmetric – similar to (**b**) – when *applied to* a low-entropy initial ensemble (such as an individual state) in a given direction of time. It would then describe the usual increase of ensemble entropy by uncontrollable 'perturbations'. The distinction between (**b**) and (**c**) requires an absolute direction of time

$\alpha \to 2\alpha \bmod 1$ on the interval $(0, 1]$. (Its α-measure would be invariant under translations in α if chosen as 'fundamental dynamics'.) The dynamical increase of this 'phase space volume' element $d\alpha$ by multiplication by the factor 2 in this map could be uniquely inverted on the *infinite* continuum, although it represents an indeterminism in the sense of the measure. However, the second term, mod 1, characterizes a semigroup as in Fig. 3.8c. Both dynamical parts are combined here in order to form a *Frobenius–Perron map*, defined everywhere on the interval in spite of representing a semigroup (symbolically indicated for the discrete case of Fig. 3.8d). The forward indeterminism, obvious in the discrete case, is often overlooked on the continuum, where the topology-conserving stretching of 'phase space' α may appear deterministic without a measure. The 'topological time asymmetry' (mod 1) contained in the Frobenius–Perron map may be phenomenologically useful, as it is able to describe the formation of macroscopic diversity. Realistic attractors must be of mixed type in order to comply with the fluctuation–dissipation theorem.

Many similar dynamical maps are discussed in the literature. They are (at most) of phenomenological value, and have little explanatory power from a fundamental statistical point of view. Their investigators often seem to regard the underlying individual microscopic reality as irrelevant. The ensembles being mapped dynamically are then treated as *real states* of physical objects. This must, of course, lead to confusion from a fundamental point of view. Statistical theories based on dynamical maps are occasionally even used for a 'minimal' interpretation of quantum mechanics (see Sect. 4.4). The misuse of purely formal ensembles as describing *physical* states is thereby reversed by identifying wave functions (that is, elementary quantum mechanical *states*) with ensembles. However, the conclusion that quantum phenomena cannot be explained in any such 'simple way' was already drawn by Bohr before the advent of matrix and wave mechanics (when his theory with Kramers and Slater had failed – see Jammer 1974).

The formation of structure is often related to a spontaneous symmetry breaking that may indeed have its origin in the fundamental *quantum indeterminism* (see Chap. 4 and Sect. 6.1). This may be the reason why the description of thermodynamical systems far from equilibrium (where structure may form) usually remains phenomenological (see Glansdorff and Prigogine 1971). The onset of structure may then be described by means of unstable fluctuations in certain quantities α, whose probabilities can be calculated from Einstein's formula (3.56). An instability would arise for them when the second derivative $\partial^2 S/\partial\alpha^2$ at a stationary point of $S(\alpha)$ becomes negative, for example by an adiabatic change in an external parameter. In this way, new *robust* quantities in the sense of Sect. 3.3.1 (see also Sect. 4.3.2) may emerge, while physical entropy is transformed into entropy of the corresponding *lacking information*, defined according to (3.58).

General Literature: Glansdorff and Prigogine 1971, Haken 1978, Cross and Hohenburg (1993).

3.5 Cosmic Probabilities and History

I shall close this chapter with a brief discussion of an objection against the probability interpretation of entropy when applied to the whole Universe and its evolution. It was first raised by Bronstein and Landau (1933), and later in a more explicit form by von Weizsäcker (1939) – see also Feynman (1965), but it may also be affected by some recent developments in cosmology.

The present state of the Universe does not only possess an entropy $S_{\alpha(\mathrm{now})}$ that is much smaller than its equilibrium value S_{equil}; it also contains documents which strongly indicate that the entropy has always been increasing during the past, $\mathrm{d}S_{\alpha(t)}/\mathrm{d}t > 0$ for $t < t_{\mathrm{today}}$. One may now compare the probability that these documents (including our private memories) have indeed formed in such a *historical* process with the probability for their formation in a mere chance fluctuation. In the former case one has

$$S_{\alpha(\mathrm{yesterday})} < S_{\alpha(\mathrm{now})} \ll S_{\mathrm{equil}} \,. \tag{3.64}$$

However, if Einstein's measure of probability in terms of entropy (3.56) were applicable to the Universe, the formation of its present state in a chance fluctuation – as improbable as it may appear – would be far more probable than a state with much lower entropy in the distant past. This objection evidently undermines Boltzmann's explanation of the thermodynamical arrow of time as arising from a grand fluctuation that occurred in an eternal universe (see also Sect. 5.3), since this fluctuation could be replaced by a far smaller one when its size is measured in terms of entropy.

This probability argument requires that the left inequality (3.64) is valid not only with respect to physical (local) entropy, but also for an appropriate ensemble entropy that takes into account all those non-local correlations which represent the convincing *consistency of documents*. Their existence in a historical universe is related to Lewis' 'overdetermination of the past' (see footnote 1 of Chap. 2).[13] While the improbability of the present solar system, for example, as having occurred in a chance fluctuation would be 'moderate' compared to that for a corresponding whole universe, the former would then have to contain consistent though unexplainable documents about the latter. David Hume's fundamental insight that we can never *predict* anything with certainty (not even that the sun will rise again) applies to the past as well – even if we did not question the general validity of the dynamical laws. Strictly speaking, we cannot be sure about the existence of *any facts* that we seem to remember. The reliability of memories and documents is in principle as doubtful as that of predictions; only the *subjective local present* cannot be questioned. Hence, even Kant's premise that we are making experience cannot be taken for granted. Not what has been observed, only our (perhaps deceiving) 'memory' that we are aware of now is beyond doubt. Saint Augustine concluded in a similar way in his *Confessiones* that the past and the future 'exist' only in the present – namely as memory and expectation 'in the soul'. This long-standing philosophical debate seems to be deeply affected (though not overcome) by thermodynamical and statistical considerations.

However, Saint Augustine's epistemologically rigorous concept of reality is obviously too restrictive for the construction of a 'world model', which must in principle always remain hypothetical (Poincaré 1902, Vaihinger 1911). The probabilistic objection raised above, even if formally correct, will thus hardly be accepted as demonstrating that causality is an illusion, based on an accident. Einstein's probabilities (3.56) for the occurrence of non-equilibrium states α, motivated by the statistical interpretation of entropy, can indeed be justified only for those macroscopic properties α which have a chance of occurring repeatedly within relevant times ('quasi-ergodically') on a generic trajectory – that is, for properties which are *not robust* on relevant timescales (hence not for stable macroscopic properties).

Physical cosmology can fortunately be derived from the more *economical* hypothesis of a universe of finite age. A homogeneous (structureless) low entropy initial state appears more acceptable in this sense than a complex state with a similarly low value of entropy. Probabilities for later states can then be calculated as *probabilities for histories* (products of successive conditioned probabilities). For example, the folding of protein chains is usually calculated along trajectories of monotonically increasing entropy (according to a master equation). Final configurations not accessible through such histories would thus be excluded even when possessing relatively large entropy. (Quantum

[13] States containing consistent (though possibly deceiving) documents were called *time-capsules* by Barbour (1994a) – see Sect. 6.2.2.

mechanically, there is always a non-vanishing but extremely small tunneling probability for their occurrence.) Most probable under the initial condition are those final states that are accessible through the most probable histories. This picture *explains* consistent documents. The thus conditioned probability for an observable world such as ours having evolved *somewhere* in the Universe would even *grow* with its size (in contrast to the global initial probability). This argument may lend support to many kinds of 'multiverses' (see Tegmark et al. 2006), which are reasonable conceptions when extrapolated from the observable universe by means of empirically founded laws or symmetries.

Whether the situation of a universe which contains scientists observing it can be regarded as probable in this sense, or whether additional 'weakly anthropic' selection criteria are required[14], has hardly ever been estimated in a reliable and unbiased way. Only at a tremendously later age of our universe could a state of maximum entropy be reached via improbable intermediate states or through quantum mechanical tunnelling (Dyson 1979), such that unconditioned probabilistic arguments apply. The cosmologically very early time that we are living at may thus remain the major improbable fact.

A 'plausible' low-entropy initial state of the Universe will be considered in Sect. 5.3. Its discussion requires quantum theory. Quantum indeterminism, whatever its correct interpretation (see Sect. 4.6), may even allow the assumption of a *unique* 'initial' state of the Universe (with a very small entropy capacity) – see Chap. 6. However, it may be worth noticing that the outcome of evolution (including ourselves) must already have been *contained as a possibility* in the huge configuration space that represents the fundamental kinematical concepts – regardless of all probability arguments.

[14] The weak anthropic principle states that we are encountering a *rare* local situation (such as a planet like Earth or a special universe in a multiverse), since we could not exist somewhere else, while the strong principle requires that the whole Universe or Multiverse must fulfill very specific conditions in order to allow our existence as observers. It has even been claimed to possess 'predictive power'. The border line between the weak and the strong principle is shifting in modern cosmology. (See Barrow and Tipler 1986, and Sect. 6.1.)