Notes on Thermodynamics and Statistical Mechanics

Wayne C. Myrvold
wmyrvold@uwo.ca
July 2013

Contents

1 Arrows of time and time-reversal symmetry

1.1 Arrows of time

Intuitively, there are many sorts of difference between the future-pointing direction of time, and the past-pointing direction. Consider

- Psychological. We remember the past, anticipate the future. There is also alleged to be a 'feeling of the passage of time.'
- Biological. We biological organisms go through life stages in a determinate order: conception (or birth by fission), embryo, youth, maturity, senescence, death.
- Thermodynamic. The laws of thermodynamics are not invariant under time reversal. Typically it is the Second Law that is said to be the culprit.
- Radiative. EM radiation is commonly found radiating outward from stars, light bulbs, burning matches, *etc.* Much more rarely, if at all, do we find spherical waves converging on an object and being absorbed by it.
- Causal. A cause comes before an effect, not *vice versa*.
- Humpty Dumpty. You can't unscramble an egg.

This is not meant to be an exhaustive list. Nor are the items in the list independent of each other; there seem to be interesting relations between them. An attractive position is that all of these are reducible to one of them (think about how this might work).

One position that has been defended is that temporal asymmetries are only apparent; to beings like us, with a temporally asymmetric perspective on the world, there *seems* to be a distinction in the physical world. I don't buy it, but see Price (1996) for an extended defense of this view.

1.2 Time-Reversal Operations

A physical theory, typically, represents the state of a system by a point in some state space Ω.

- The state of a single classical particle is represented by its position and momentum $(\mathbf{x}, \mathbf{p})$, which can be thought of as a point in its 6-dimensional *phase space*.
- The state of n classical particles is represented by a point in a $6n$-dimensional phase space.
- The state of a quantum system is represented by a vector in a complex Hilbert space.
- The *thermodynamic* state of a system is given by a small (compared to the dimension of the phase space of all the particles that make up the system) number of macroscopically measurable parameters.

A state history is a trajectory through phase space; that is, a mapping $\sigma : I \subseteq \mathbb{R} \to \Omega$, for some time interval I.

A physical theory will also typically include a set of dynamical laws that distinguish, from among the kinematically possible trajectories, a set $\mathcal{D}$ of dynamically possible trajectories.

Given any time t_0, we can define a reflection of the time axis around t_0 by

$$t \to t^T = t_0 - (t - t_0) \tag{1.1}$$

It's traditional to take $t_0 = 0$, so that $t^T = -t$.

We can also talk about time-reversal of states.

1.2.1 Ex. 1. Classical particle

For a classical particle, the operation of time-reversal is (perhaps obviously)

$$(\mathbf{x}, \mathbf{p}) \to (\mathbf{x}, -\mathbf{p}). \tag{1.2}$$

1.2.2 Ex. 2 EM

In electromagnetic theory the standard view of time-reversal has charges remaining invariant, velocities (and hence currents) changing sign.

$$\begin{aligned} \rho &\to \rho \\ \mathbf{J} &\to -\mathbf{J}, \end{aligned} \tag{1.3}$$

with fields going as

$$\begin{aligned} \mathbf{E} &\to \mathbf{E} \\ \mathbf{B} &\to -\mathbf{B} \end{aligned} \tag{1.4}$$

There is a heterodox view, on which the time-reversal operation should leave *both* electric and magnetic fields invariant. See Albert (2000). For another heterodox view, on which time-reversal operation flips the sign of charges as well, see Arntzenius and Greaves (2007) and Leeds (2006).

For two, somewhat different, defenses of the orthodox view, see Earman (2002) and Malament (2004).

1.2.3 Ex. 3. QM

In quantum mechanics, the state of a spinless particle is represented by a wave-function ψ.

To see how to time-reverse this, consider the following:

1. By Wigner's theorem,[1] there exists a either a unitary transformation or an antiunitary transformation that implements the operation.

2. We want

   (a) $\langle \hat{x} \rangle \to \langle \hat{x} \rangle$

   (b) $\langle \hat{p} \rangle \to -\langle \hat{p} \rangle$

This yields the result that the wave function's transformation under time reversal is given by the anti-unitary transformation

$$\psi \to \psi^*. \tag{1.5}$$

When spin is to be taken into account, the time-reversal operator is defined so as to flip the signs of spins. (Rationale: spin, being a form of angular momentum, should change sign, just as orbital angular momentum does.)

For a heterodox view, according to which time-reversal should leave ψ unchanged, see Callender (2000).

[1] See Weinberg (1995, Ch. 2, Appendix A) for an exposition.

1.3 Time-Reversal Invariance of dynamical laws

.

We are now ready to consider time-reversal invariance of physical laws. Given a time-reflection

$$t \to t^T = t_0 - (t - t_0) \tag{1.6}$$

and a state-reversal operation

$$\omega \to \omega^T, \tag{1.7}$$

we can define an operation that reverses state histories. Define the history-reversal operation

$$\sigma \to \sigma^T \tag{1.8}$$

by

$$\sigma^T(t) = \sigma(t^T)^T. \tag{1.9}$$

So, if σ includes a sequence of states $...\sigma(t_1), \sigma(t_2), \sigma(t_3)...$, then the time-reversed history includes a sequence of states $...\sigma(t_3^T)^T, \sigma(t_2^T)^T, \sigma(t_1^T)^T, ...$

We say that a physical theory is *time-reversal invariant* iff, whenever a state history σ is dynamically possible, the time-reversed state history σ^T is too. Or, in other words, the theory is time-reversal invariant iff $\mathcal{D}^T \subseteq \mathcal{D}$.

1.3.1 Example. Newtonian mechanics.

Suppose we have a system of n Newtonian particles. Newton's 2nd law says that

$$m_i \frac{d^2\mathbf{x}_i}{dt^2} = \mathbf{F}_i, \tag{1.10}$$

where $\mathbf{F}_i$ is the total force on the ith particle.

Given a system of Newtonian particles, suppose that the force on any one of the particles depends only on the positions of the particles.

$$\mathbf{F}_i = \mathbf{F}_i(\mathbf{x}_1, ..., \mathbf{x}_n). \tag{1.11}$$

Then both left and right-hand sides of Equation (1.10) remain unchanged under time-reversal, and so the law of motion is time-reversal invariant.

If, however, the force depends on velocity, then we will *not* have time-reversal invariance. Consider, *e.g.* a damped harmonic oscillator, whose equation of motion is

$$m\,\ddot{x} = -kx - b\dot{x}. \tag{1.12}$$

If $x(t)$ is a solution to this equation, then, except for the trivial (equilibrium) solution $x(t) = 0$, its time reversal will not be a solution. Solutions to Equation (1.12) are oscillations with exponentially decreasing amplitude (assuming subcritical damping). The time-reverse of such a solution would be an oscillation with exponentially *increasing* amplitude.

(To ponder: in damping, what is going on at the microphysical level?)

1.3.2 EM

Maxwell's equations are

$$\begin{aligned} \nabla \cdot \mathbf{E} &= 4\pi\rho & \nabla \times \mathbf{E} + \tfrac{1}{c}\tfrac{\partial \mathbf{B}}{\partial t} &= 0 \\ \nabla \cdot \mathbf{B} &= 0 & \nabla \times \mathbf{B} - \tfrac{1}{c}\tfrac{\partial \mathbf{E}}{\partial t} &= \tfrac{4\pi}{c}\mathbf{J} \end{aligned} \tag{1.13}$$

Under time reversal

$$\begin{aligned} \nabla &\rightarrow \nabla \\ t &\rightarrow -t \\ \rho &\rightarrow \rho \\ \mathbf{J} &\rightarrow -\mathbf{J}, \\ \mathbf{E} &\rightarrow \mathbf{E} \\ \mathbf{B} &\rightarrow -\mathbf{B} \end{aligned}$$

And so we have TRI.

1.3.3 Example. QM

The law of motion is the Schrödinger equation:

$$i\hbar\,\frac{\partial}{\partial t}\Psi = \hat{H}\Psi. \tag{1.14}$$

Suppose the Hamiltonian for a particle takes the form:

$$\hat{H} = -\frac{\hbar^2}{2m}\nabla^2 + V(\hat{\mathbf{x}}). \tag{1.15}$$

Then, if $\Psi(\mathbf{x}, t)$ satisfies the S. eq., then $\Psi^T(\mathbf{x}, t) = \Psi(\mathbf{x}, -t)^*$ satisfies

$$-i\hbar \frac{\partial}{\partial t}\Psi^T = \hat{H}\Psi^T, \tag{1.16}$$

or,

$$i\hbar \frac{\partial}{\partial t^T}\Psi^T = \hat{H}\Psi^T, \tag{1.17}$$

which is the time-reversed Schrödinger equation. So QM is TRI provided that the Hamiltonian is invariant under time-reversal.

1.3.4 Example. Weak-force interactions

At this point, you may be forming the induction: physical laws are, at the fundamental level, time-reversal invariant, and perhaps conjecturing that, for deep reasons, they *must* be. Before dashing off an *a priori* proof of this, pause to consider: there is evidence that interactions involving the weak nuclear force are *not* TRI. Indirect evidence of this first came through evidence of CP violation; by the CPT theorem, the same processes must involve T-violation. The first direct evidence of violation of T-symmetry, involving neutral kaon decays, was reported in Angelopoulos et al. (1998). T-violation has also been recently observed in the B-meson system (Lees et al., 2012).

Thus, it can't be a metaphysical necessity that the laws of physics are TRI. However, there are principled reasons for regarding these violations of time reversal invariance as irrelevant to the temporal asymmetries associated with thermodynamics. One way to see this is that, although the laws of nature are invariant under time reversal, there is good reason to think (as this follows from Lorentz invariance) that they *are* invariant under CPT: the operation that combines charge conjugation (particle $\rightarrow$ antiparticle), parity (that is, mirror) reflection, and time reversal. Unless a gas consisting of anti-hydrogen molecules exhibits anti-thermodynamic behaviour, the temporal asymmetries of thermodynamics are also violations of CPT symmetry.[2]

1.4 Is Time-Reversal Invariance Trivial?

If you take a look at textbook accounts of the treatment of time-reversal operations in EM, or in QM, they fall into two classes. There are those that justify the standard time-reversal operation on the grounds that it is what it

[2]This point has been made by Price (2006).

takes to make the theory TRI, and those that attempt to give an independent rationale for the operation. The former strategy engenders the suspicion that we are cooking up the time-reversal operation to save a symmetry that we happen to like. This charge has been laid, for EM, by Albert (2000), chapter 1, and, for QM, by Callender (2000). And one distinguished author has argued that, for *any* deterministic theory, it is possible to cook up a time reversal operation that the theory is TRI.

> If one were to allow *completely arbitrary* time reversal operations, then any deterministic theory would count as time reversible. For example, take some particular state history $S(t)$ that is possible according to the theory. Let us now define what the time inverse S^T is of any state that lies on that particular history. Begin by arbitrarily choosing some time t_0. Then declare that the time reversal of the state $S(t_0 + \delta t)$ that occurs a period of time δt after t_0 is, in fact, the state that occurs time period δt before t_0, i.e., $S(t_0 - \delta t)$.
>
> ... given any deterministic theory one can define a time reversal operation T that shows that the theory in question is time reversible. But that is absurd. (Arntzenius, 2004, 32–33)

Arntzenius is right that it is absurd. But it's worthwhile to think about *why* it's absurd. (How would this sort of "time reversal operation" look when applied to, say, a damped harmonic oscillator, which exhibits *observable* temporal asymmetry?)

A theory that exhibits temporal asymmetry at the level of observable phenomena is clearly not a candidate for a theory that is TRI. Suppose, then, that the observable phenomena do not pick out a distinguished direction of time — for any sequence of observations, the reversed sequence is possible, according to the theory. (Remember, we do not observe EM fields directly, but only through their effects on charged particles). If, now, we held a positivist view, according to which the observable consequences of a theory exhaust its physical content, and any apparent reference to structure that is not directly observable merely 'dressing' of the real content of the theory, we'd be done, and declare the theory TRI. There are good reasons (beyond the scope of this course) for not adopting such a narrow view of the content of a theory. However, if, according the theory, the phenomena fail to distinguish temporal orientations, this at least *suggests* that the theory is TRI, and that

apparent temporal asymmetries in our theoretical treatment are artifacts of our representation. We ought, therefore, to ask ourselves whether, by reflecting on the physical significance of the part of the theory that refers to structures that are not directly observable, we will conclude that the appropriate time-reversal operation shows the theory to be TRI.

That is, we are imagining something analogous to Einstein's train of thought in his 1905 paper, "On the Electrodynamics of Moving Bodies." Einstein begins with the observation that electromagnetic phenomena depend, not on the absolute motions of the systems involved, but only on their relative motions. Our theoretical description, on the other hand, introduces an asymmetry that is not present in the phenomena. This at least suggests that the asymmetry is an artifact of our representation. Einstein then shows that it is possible to formulate the theory in a way that does *not* distinguish between states of inertial motion; on this formulation the partition of the electromagnetic force on a charged body into electrical and magnetic components is explicitly a frame-relative matter.

Malament (2004) does something analogous for time reversal in EM. If we consider a system of charges moving under the influence of electromagnetic forces, their behaviour does not distinguish between past and future directions of time. This suggests that we try to formulate the theory in a way that does not distinguish the two directions. The key to how to do this comes from the following remark:

> We can think of it [the EM field tensor] as coding, for every point in spacetime, the electromagnetic force that would be experienced by a point test particle there, depending on its charge and instantaneous velocity.

Malament shows that it is, in fact, possible to provide a coordinate-free representation of the theory that does not require a temporal orientation — that is, does not require us to declare which of the two temporal direction is past, and which is future. It turns out that, on this formulation, components of the electromagnetic field tensor are defined only given a choice of temporal orientation and handedness of spatial coordinate system. Malament's treatment leads naturally to the conclusion that, under temporal inversion,

$$\begin{aligned} \mathbf{E} &\to \mathbf{E} \\ \mathbf{B} &\to -\mathbf{B}. \end{aligned} \tag{1.18}$$

The upshot of the analysis is that the magnetic field is represented by an *axial vector*.

> The transformation properties for B^a are exactly the same as for angular velocity. ... Magnetic field might not be rates of change of anything in the appropriate sense, but they *are* axial vector fields (Malament, 2004, 313–314).

This is not news; Maxwell already knew this!

> The consideration of the action of magnetism on polarized light leads, as we have seen, to the conclusion that in a medium under magnetic force something belonging to the same mathematical class as an angular velocity, whose axis in the direction of the magnetic force, forms a part of the phenomenon. (Maxwell, 1954, §822).

Maxwell went on to argue that magnetic fields were, in fact, associated with vortices in the electromagnetic ether. We have abandoned the ether, but the conclusion that magnetic fields transform, under parity and time reversal, in the same way that an angular velocity does, survives the abandonment.

2 Basic Concepts and Laws of Thermodynamics

Early writers on thermodynamics tended to talk of two fundamental laws of thermodynamics, which came to be known as the First and Second Laws of Thermodynamics. However, another law has been recognized, regarded as more fundamental than the first two, which has accordingly come to be known as the Zeroth Law. But perhaps there is also a Minus First Law, more fundamental than all of these (see Brown and Uffink (2001) for extended discussion). There is also a Third Law of Thermodynamics (which, though important, will play less of a role in our discussions).

2.-1 The Minus First Law and Thermodynamic Equilibrium

> Thermodynamicists get very excited, or at least get very interested, when nothing happens... . (Atkins, 2007, p. 7)

On the first page of Pauli's lectures on thermodynamics we find,

> Experiment shows that if a system is closed, then heat is exchanged within the system until a stable thermal state is reached; this state is known as *thermodynamic equilibrium.* (Pauli, 1973, p. 1)

Similar statements can be found in the writings of others, and, even when not explicitly stated, it is taken for granted. Brown and Uffink have dubbed this the Minus First Law, or *Equilibrium Principle*, which they state as,

> *An isolated system in an arbitrary initial state within a finite fixed volume will spontaneously attain a unique state of equilibrium.* (Brown and Uffink, 2001, p. 528)

Note that this is a time-asymmetric law. Once an isolated system attains equilibrium, it never leaves it without outside intervention; the time reversal of this is not true. Brown and Uffink argue that it is this law, not, as most writers on the subject would have it, the Second, that is at the heart of the time asymmetry of thermodynamics.

It is interesting that time asymmetry is present in what is perhaps the most fundamental concept of thermodynamics, that of equilibrium.

2.0 The Zeroth Law of Thermodynamics

The concept of thermodynamic equilibrium can be used to introduce the concept of temperature. Two bodies that are in thermal contact (this means that heat flow between them is possible), which are in thermal equilibrium with each other, will be said to have the *same temperature*. We want this relation of equitemperature to be an equivalence relation. It is reflexive and symmetric by construction. That it is transitive is a substantive assumption[3] (though one that has often been taken for granted). Suppose we have bodies that can be moved around and brought into thermal contact with each other. When this happens, the contact might induce a change of state (brought on by heat transfer from one to the other), or it might not. The zeroth law says that, if two bodies A, B are in equilibrium with each other when in thermal contact, and B and C are in equilibrium with each other, then A and C are in equilibrium with each other.

2.0.1 Ideal gases and thermometry

A recurring example we will use will be *ideal gases*. An ideal gas has a particularly simple thermodynamic state space: its equilibrium thermodynamic states are determined by the pressure, temperature and volume of the gas, and, because these are related by the *equation of state*, there are only two independent parameters, so we have a two-dimensional state space.

An ideal gas is defined to be one satisfying

- *Boyle's Law.* At fixed temperature,

$$p \propto \frac{1}{V}. \tag{2.1}$$

- *Joule's Law.* The internal energy depends only on the temperature.

Note that these depend only on the notion of *same temperature*, introduced on the basis of the 0th Law. Both of these are obeyed to a good approximation by real gases, provided that the density is not too high.

Before we can write the equation of state down, we need to introduce a temperature scale.

Boyle's law entails that there is a function of the thermodynamic state of the gas, call it $\theta \propto pV$, that takes on the same value at equal temperatures

[3] Noted and emphasized by, among others, Maxwell; see Maxwell (2001, pp. 32–33)

(note that the Zeroth law gives us a right to talk about equality of temperature, even in the absence of a quantitative measure). This gives us, for any two states of the gas,

$$\frac{\theta_1}{\theta_2} = \frac{p_1 V_1}{p_2 V_2} \tag{2.2}$$

and hence, a numerical 'ideal gas temperature' defined up to an arbitrary scale factor. Note that this definition of temperature includes a non-arbitrary zero point: the temperature that is approached as the volume of our gas goes to zero. Choose a standard temperature and standard pressure (STP). This is arbitrary, but it is conventional to choose 0°C and 100 kPa. Our gas will have volume $V_s = V(\theta_s, p_s)$ at STP, and we can rewrite (2.2) as,

$$pV = \left(\frac{p_s V_s}{\theta_s}\right)\theta. \tag{2.3}$$

Different samples of an ideal gas will have different values for the factor in equation (2.3); intuitively, the volume of a gas at STP depends on how much 'stuff' is in it. Pick a reference volume V_0, and define,

$$n = V_s(\theta_s, p_s)/V_0. \tag{2.4}$$

That is, n is how many multiples there are in our sample of the amount of gas that would occupy the reference volume V_0 at STP. If we choose V_0 to be the standard molar volume, approximately 22.4 liters, then n will be the number of moles of gas in our sample.[4] Then we have

$$pV = n\left(\frac{p_s V_0}{\theta_s}\right)\theta. \tag{2.5}$$

The quantity in parentheses is now purely a scale factor, dependent only on our choice of units. Call it R.[5] Then we have

$$pV = nR\,\theta. \tag{2.6}$$

Note that this is dependent only on Boyle's law; the parameter θ is defined to be the quantity measured by an ideal-gas thermometer. In §2.2.6 we will

[4]Note, however, that nothing we have said so far has committed us to a gas being composed of discrete molecules, and everything in this section would make sense if a gas were a continuous fluid.

[5]With θ_s, p_s chosen to be STP, and V_0 chosen to be the standard molar volume, R is the usual ideal gas constant.

introduce, via the Second Law, a notion of *thermodynamic temperature* T, which will turn out to be proportional to θ. Choosing equal units for θ and T gives us the ideal gas equation of state in its familiar form,

$$pV = nRT, \tag{2.7}$$

which is called the *ideal gas law*.

2.1 The First Law: Heat and Work

We are used to the idea that energy is conserved. A system of bodies will have a total internal energy that consists, in part, of the kinetic energy of its components, which may be in motion, and in part, to the potential energy due to the forces acting between them.

One way to add energy to a system is to *do work* on it. For example, I may compress a spring. The compressed spring has a potential energy, which can be converted to kinetic energy. Or I can lift a weight, which gains gravitational potential energy, which again can be recovered as kinetic energy. The energy that I can get back out, when I compress an ideal spring from an initial state S_i to a final state S_f, is equal to

$$W = -\int_{S_i}^{S_f} \mathbf{F} \cdot d\mathbf{x}, \tag{2.8}$$

where $\mathbf{F}$ is the force opposing my efforts. It is defined this way so that the increase of energy of the system on which I do work is equal to the work I do.

Suppose, now, that I expend the same amount of work by, say, stirring a viscous fluid, or by rubbing two rough surfaces together. I won't be able to recover as kinetic energy the energy I put in as work — at least not all of it. The system I did work on, however, will get warmer. The energy I expended did not vanish; it was converted into heat.

Here, again, it might seem like we're cheating. Every time there's an apparent violation of conservation of energy, we invent a new form of energy — first potential energy, and then heat — to compensate for the apparently lost energy. This might make the principle of conservation of energy seem to be an empty one — with enough imagination, it might seem, we could come up with a new form of energy to make energy conservation true, no matter what happens. However, what gives the notion some teeth is the fact

that there is a measurable *mechanical equivalent of heat*: we can measure the amount of work it requires to raise, say, a gram of water 1 degree.

The First Law of Thermodynamics says that, if an amount of work W is done on a system, and heat Q passes into it, the internal energy U of the system is changed by an amount

$$\Delta U = Q + W \tag{2.9}$$

Note: the terms *work* and *heat* are used in connection with two modes of energy transfer. I can do work on a system, and transfer energy to it that way. Energy can also be transferred as heat flow between two bodies in thermal contact. We're tempted to think of heat as a substance that can flow from one body to another (indeed, this was at one time a theory that was taken seriously), but on the modern view, energy transferred as work can be extracted as heat, and *vice versa*, and it makes no sense to talk about the heat content of a body.

We will often want to integrate the change of internal energy along some process. For that reason, the differential form of the First Law will frequently be more useful.

$$dU = đQ + đW \tag{2.10}$$

Writing 'dU' indicates that the quantity is a change in a function of state: some quantity that depends only on the thermodynamic state of the system. As mentioned above, we do not ascribe to a body some quantity Q that represents the heat it contains. The small heat transferred is not a change in a state function, and we write it as '$đQ$' to signal this. We call a quantity such as dU, that is a change in a state function, an *exact differential*, and a quantity, such as $đQ$ or $đW$, that is not a change in a state function, and *inexact differential*.

A bit of jargon: a system is *adiabatically isolated* iff it can't exchange heat with the environment, and an *adiabatic process* is one in which the system exchanges no heat with the environment.

2.1.1 Heat capacity

Define the *constant-volume heat capacity* of a gas as the amount of heat $đQ$ required to raise the temperature by an amount $d\theta$.

$$C_V = \left(\frac{đQ}{d\theta}\right)_V = \left(\frac{\partial U}{\partial \theta}\right)_V, \tag{2.11}$$

where the subscript V indicates that the heating is being done at constant volume. We can also heat a gas at constant pressure, say, by allowing it to raise a piston with a weight on it, and define *constant-pressure heat capacity* as

$$C_p = \left(\frac{đ Q}{d\theta} \right)_p . \tag{2.12}$$

In constant-pressure heating, some of the energy I put in goes into raising the temperature, and some into expansion (the gas does work on the environment). We should, therefore, expect that it takes more heat to raise the temperature of a gas one degree under constant pressure than it does under constant volume, that is,

$$C_p > C_V. \tag{2.13}$$

For the quantitative relations between heat capacities, see section 2.2.3, below.

2.2 The 2nd law and Entropy

> But by reason of the Tenacity of Fluids, and Attrition of their Parts, and the Weakness of Elasticity in Solids, Motion is more apt to be lost than got, and is always upon the Decay.
>
> Newton, *Opticks* (Newton, 1952, p. 398)

2.2.1 Quasistatic, reversible processes

A central distinction in thermodynamics is between two kind of processes. On the one hand, there are processes that take place gently, with no churning or turbulence or friction, and which no heat is transferred from a warmer body to a colder (all of these things involve lost opportunities to do work with the energy transferred). On the other hand, there are all other processes.

Example: if I compress a gas, then at minimum I have to exert a force on the gas that is equal to the opposing pressure. But if the force I exert is exactly the same, nothing happens. However, assuming a frictionless piston, any slight push I make, above and beyond the force needed to hold the piston in place, will compress the gas, and, provided I am willing to wait long enough, I can compress the gas using arbitrarily small force beyond that which counteracts the pressure.

The word 'quasistatic' is often used in connection with such processes, as is the word 'reversible.' A quasistatic process is meant to be one that is carried out so slowly that, at every moment, the system is effectively in equilibrium. Reversibility is meant to indicate that the initial state is recoverable. If I compress a gas slowly by doing work on it, say, by allowing a spring to extend, then I can get the energy transferred out by allowing the gas to expand and recompress the spring. Note that 'reversible' here doesn't necessarily mean that the time-reverse of the process is possible. For a careful discussion of these concepts, see Uffink (2001).

What we want are processes that are quasistatic *and* reversible. I will usually say 'qsr.' However, there doesn't seem to be a good adverbial form of this, so I will sometimes say 'quasistatically' when what I really mean is, 'in a qsr manner.'

2.2.2 Work done on a gas

Suppose we compress a gas quasistatically by an amount dV, by moving a piston of area A a distance dx. Since this is a compression, the volume decreases.

$$dV = -Adx \tag{2.14}$$

The force opposing this compression is due to the pressure exerted by the gas by the piston. Since pressure is force per unit area, the force exerted on the piston is pA , and the work I do on the has in compressing it is

$$đW = pAdx = -pdV. \tag{2.15}$$

Hence, a useful form of the First Law for a gas (or any system that can only do work on the outside world by expanding) is that, for any qsr process,

$$dU = đQ - p\,dV. \tag{2.16}$$

2.2.3 Relations between heat capacities

From eq. 2.16 it follows that, for any gas,

$$\left(\frac{\partial U}{\partial \theta}\right)_p = \left(\frac{đQ}{d\theta}\right)_p - p\left(\frac{\partial V}{\partial \theta}\right)_p, \tag{2.17}$$

and so,

$$C_p = \left(\frac{đQ}{d\theta}\right)_p = \left(\frac{\partial U}{\partial \theta}\right)_p + p\left(\frac{\partial V}{d\theta}\right)_p. \tag{2.18}$$

It can also be shown (left as an exercise for the reader), that, for any gas,

$$C_p - C_V = \left[p + \left(\frac{\partial U}{\partial V}\right)_\theta\right]\left(\frac{\partial V}{\partial \theta}\right)_p. \tag{2.19}$$

Joule's law says that the internal energy of an ideal gas depends only on its temperature, and hence,

$$\left(\frac{\partial U}{\partial V}\right)_\theta = 0. \tag{2.20}$$

From the ideal gas law we have,

$$\left(\frac{\partial V}{\partial \theta}\right)_p = \frac{nR}{p}, \tag{2.21}$$

and so, for an ideal gas,

$$C_p - C_V = nR. \tag{2.22}$$

It is convenient to work with the *molar heat capacities* $c_p = C_p/n$, $c_V = C_V/n$, related by

$$c_p - c_V = R. \tag{2.23}$$

We also define

$$\gamma = \frac{C_P}{C_V} = \frac{c_p}{c_V}. \tag{2.24}$$

Note that, given the way we have defined heat capacities, they could, in principle (and for some systems do) vary with the state of the system. However, experience shows that the heat capacity of an ideal gas does not change with temperature. This, together with Joule's law, gives us,

$$dU = C_V\, d\theta. \tag{2.25}$$

Since left and right side are state functions, this is true for *any* change of state, whether reversible or not. It also follows that, since the difference between C_P and C_V is a constant, that C_P, and hence γ, is also constant.

2.2.4 A useful relation

For an ideal gas undergoing a qsr, adiabatic process,

$$dU = C_V d\theta = -pdV \tag{2.26}$$

From the ideal gas law, $p = nR\theta/V$, and so we have

$$\frac{d\theta}{\theta} = -\left(\frac{R}{c_v}\right)\frac{dV}{V} = -(\gamma - 1)\frac{dV}{V}. \tag{2.27}$$

Integrating this gives the conclusion that, for an adiabatic, qsr process

$$\theta\, V^{\gamma-1} = const. \tag{2.28}$$

2.2.5 Carnot's theorem

Consider a heat engine that absorbs heat Q_{in} from a heat reservoir, does net work W on the external world, and discards some waste heat Q_{out} into another (cooler) reservoir. (These reservoirs are to be regarded as so large that they can supply or absorb these quantities of heat with negligible change in temperature.) Suppose, further, that the heat engine operates in a cycle, so that it returns to its original thermodynamic state at the end of the cycle. This means that the engine undergoes no net change in internal energy. Conservation of energy entails

$$W = Q_{in} - Q_{out}, \tag{2.29}$$

The engine absorbs heat Q_{in} from the hot reservoir, and converts fraction

$$\eta = \frac{W}{Q_{in}} \tag{2.30}$$

of it into useful work, and discards the rest. The fraction η is called the *efficiency* of the engine.

$$\eta = 1 - \frac{Q_{out}}{Q_{in}}. \tag{2.31}$$

Carnot's theorem tells us about the maximum efficiency of such an engine:

> Any two heat engines operating in a qsr manner between two heat reservoirs have the same efficiency, which is dependent only on the temperature of the two reservoirs. Moreover, any *other* heat engine has lower efficiency.

The argument for this is based on the observation that, though heat flows spontaneously from a hot to a cold body, we have to *do something* – *e.g.* expend some work — to transfer heat from a cold to a hot body. Clausius expressed the latter idea by,

> Heat cannot pass from a colder body to a warmer body without some other change connected with it occurring at the same time.[6]

This is often called the *Clausius statement of the Second Law of Thermodynamics.*

Here's the argument. A given reversible engine $\mathcal{E}_r$, run forward, extracts heat from the hot reservoir and converts a portion of it into work. The cycle can be run backward: If we do work on $\mathcal{E}_r$, we can use it to move heat from the cold reservoir to the hot; that is, we use it as a refrigerator. Let the efficiency of $\mathcal{E}_r$ be η_r, and let $\mathcal{E}_s$ be some other engine, with efficiency η_s.

Let $\mathcal{E}_s$ extract an amount of heat Q_H from the hot reservoir, do work $W = \eta_s Q_H$ on $\mathcal{E}_r$, and discard heat $Q_C = (1 - \eta_s)Q_H$. The work done on $\mathcal{E}_r$ is used to drive it backwards, extracting heat Q'_C from the cold reservoir, and dumping heat Q'_H into hot reservoir. We have,

$$W = \eta_s Q_H = \eta_r Q'_H, \tag{2.32}$$

and so, the net result of the cycle is to move a quantity of heat

$$Q = Q'_H - Q_H = \left(\frac{\eta_s}{\eta_r} - 1\right) Q_H \tag{2.33}$$

from the cold reservoir to the hot reservoir. If $\eta_s > \eta_r$, this is positive, and the net result of the process was to move heat from the cold reservoir to the hot, which L2 says is impossible. Conclusion:

> For engine $\mathcal{E}_s$ and any reversible engine $\mathcal{E}_r$, $\eta_s \leq \eta_r$.

From which it follows

> All reversible engines have the same efficiency.

2.2.6 Thermodynamic Temperature

Forget, for the moment, that we have already introduced the ideal gas temperature θ. Carnot's theorem tells us that the efficiency of a reversible heat engine operating between two reservoirs depends only on their temperature;

[6] "Es kann nie Wärme aus einem kälteren Körper übergehen, wenn nicht gleichzeitig eine andere damit zusammenhängende Aenderung eintritt." Quoted by Uffink (2001, p. 333).

here the notion of 'same temperature' that we are helping ourselves to is the equivalence relation underwritten by the Zeroth Law. We can use this fact to *define* a temperature scale. If η_{AB} is the efficiency of a reversible engine operating between reservoirs A, B, define the *thermodynamic temperature* T by

$$\frac{T_B}{T_A} =_{df} 1 - \eta_{AB}. \tag{2.34}$$

This defines the thermodynamic temperature of any reservoir up to an arbitrary scale factor. With the scale chosen so that degrees are equal in size to Centigrade degrees, this is (of course) called the Kelvin scale.

2.2.7 The Carnot cycle

> For the purposes of scientific illustration, and for obtaining clear views of the dynamical theory of heat, we shall describe the working of an engine of a species entirely imaginary—one which it is impossible to construct, but very easy to understand.
>
> (Maxwell, 2001, pp. 138–139)

We now have two temperature scales on our hands: the ideal gas temperature θ, and the thermodynamic temperature T, and we may justly ask whether there is any relation between them. To this end, we will imagine a heat engine whose working substance is an ideal gas, and consider a reversible cycle that is particularly simple to analyze, get the efficiency of this cycle in terms of the ideal gas temperature, and hence get a relation between ideal gas and thermodynamic temperature scales.

Since transfer of heat between bodies of unequal temperature is an irreversible process, we will arrange our cycle so that any heat exchange occurs at constant temperature, in contact with one of the reservoirs. The cycle will be broken into four steps:

1. $a \to b$. Constant temperature expansion at temperature θ_H. Heat Q_{in} absorbed by the system from the hot reservoir. Work is done by the gas on the environment.

2. $b \to c$. Adiabatic expansion. The gas does work on the environment, cooling as it does to temperature θ_C.

3. $c \to d$. Constant temperature compression at temperature θ_C. An external agent does work on the gas.

4. $d \to a$. Adiabatic compression. Again, an external agent does work on the gas.

We want to find out what the relation is between Q_{in} and Q_{out}. What makes this easy to analyze is the fact (Joule's law again) that the internal energy of a gas is a function only of its temperature. Therefore, for the isothermal process 1,

$$\Delta U = Q_{in} - \int_a^b p\, dV = 0. \tag{2.35}$$

From the ideal gas law, $p = nRV/\theta$, and so

$$\int_a^b p\, dV = nR\,\theta_H \int_a^b \frac{dV}{V} = nR\,\theta_H \log\left(\frac{V_b}{V_a}\right), \tag{2.36}$$

giving us

$$Q_{in} = nR\,\theta_H \log\left(\frac{V_b}{V_a}\right). \tag{2.37}$$

Similarly,

$$Q_{out} = nR\,\theta_C \log\left(\frac{V_c}{V_d}\right). \tag{2.38}$$

Therefore,

$$\frac{Q_{in}}{\theta_H} - \frac{Q_{out}}{\theta_C} = nR \log\left(\frac{V_b\,V_d}{V_a\,V_c}\right). \tag{2.39}$$

Here's where the useful fact (2.28) comes in. It gives us

$$\left(\frac{V_b}{V_c}\right)^{\gamma-1} = \frac{\theta_H}{\theta_C} = \left(\frac{V_a}{V_d}\right)^{\gamma-1}, \tag{2.40}$$

which gives us

$$\frac{V_b\,V_d}{V_a\,V_c} = 1, \tag{2.41}$$

and so

$$\frac{Q_{in}}{\theta_H} = \frac{Q_{out}}{\theta_C}. \tag{2.42}$$

Therefore, the efficiency of the Carnot engine is

$$\eta_{HC} = 1 - \frac{\theta_C}{\theta_H}. \tag{2.43}$$

Comparison of (2.34) and (2.43) yields the happy result that

$$\theta \propto T, \tag{2.44}$$

and of course, the easiest convention to adopt is to use the same size units for each, in which case we have equality of θ and T. Henceforth, we will speak only of T.

2.3 Enter Entropy

Around a Carnot cycle,

$$\oint \frac{đQ}{T} = \frac{Q_{in}}{T_H} - \frac{Q_{out}}{T_C} = 0. \tag{2.45}$$

Moreover, this must be true around *any* qsr cycle in the ideal gas state space. The argument: any cycle in the state space of our system can be approximated as closely as we want by a path that alternates between isothermal and adiabatic segments, and, for such paths, (2.45) holds.

So, we conclude, for *any* thermodynamic system,

$$\oint_{qsr} \frac{đQ}{T} = 0. \tag{2.46}$$

Argument: if it didn't hold, we could construct a reversible engine with an efficiency different from the Carnot efficiency, in contravention of Carnot's theorem.

It follows that there exists a state function S such that, for any qsr process,

$$\int_a^b \frac{đQ}{T} = S_b - S_a, \tag{2.47}$$

or, in differential form,

$$dS = \left(\frac{đQ}{T}\right)_{qsr}. \tag{2.48}$$

This state function is called the *thermodynamic entropy* of the system. Note that it is defined only up to an arbitrary additive constant; it is entropy *differences* between thermodynamic states that are physically significant.

A heat engine operating between two reservoirs that is less efficient than a Carnot engine will have

$$\frac{Q_{out}}{Q_{in}} > \frac{T_C}{T_H}, \tag{2.49}$$

hence, for the cycle of such an engine,

$$\oint \frac{đQ}{T} < 0. \tag{2.50}$$

We can express the content of (2.45) & (2.50) in differential form: for any process,

$$đQ \leq T\, dS, \tag{2.51}$$

with equality holding for reversible processes. It follows that, for an adiabatically isolated system, which cannot exchange heat with the rest of the world,

$$dS \geq 0. \tag{2.52}$$

The entropy of an adiabatically isolated system cannot decrease.

2.4 Entropy of an ideal gas

We know that the entropy of any system is a function of its thermodynamic state. It will be useful to explicitly exhibit the dependence of the entropy of an ideal on its state.

Let us take an ideal gas from a state (p_a, V_a, T_a) to a state (p_b, V_b, T_b), and ask what the change of entropy is.

This will be the integral of dQ/T along any qsr path that joins the states. We have, from the First Law,

$$dU = C_V dT = đQ + đW = đQ - pdV, \tag{2.53}$$

or,

$$dS = \frac{đQ}{T} = C_V \frac{dT}{T} + \frac{p\, dV}{T}. \tag{2.54}$$

From the ideal gas law, $p/T = nR/V$, and so

$$dS = C_V \frac{dT}{T} + nR\frac{dV}{V}. \tag{2.55}$$

This yields

$$\begin{aligned} \Delta S &= C_V \log\left(\frac{T_b}{T_a}\right) + nR\log\left(\frac{V_b}{V_a}\right) \\ &= C_V \log\left(\frac{T_b\, V_b^{\gamma-1}}{T_a\, V_a^{\gamma-1}}\right). \end{aligned} \tag{2.56}$$

2.4.1 Two Examples

Free expansion. An ideal gas is in an adiabatically isolated container. It is initially confined to a subvolume V_i. A partition is removed, and the gas expands adiabatically and without doing any work to fill the volume V_f now available to it. What is the entropy increase?
Answer. Adiabatic isolation means no heat exchange with environment, and, since there was no work done either, the internal energy of the gas, and hence its temperature, is unchanged. The change in entropy is therefore

$$\Delta S = nR\log\left(\frac{V_f}{V_i}\right), \tag{2.57}$$

which is, of course, positive, as must be the case for any spontaneous process.

Diffusion. Two ideal gases, both having initial temperature T and pressure p, are initially confined to compartments of volume V_1 and V_2, respectively. The partition separating them is removed, and they diffuse into each other's compartments, coming to a new equilibrium with each of them equally distributed throughout the total volume V_f. Has the entropy increased?
Answer. One might be tempted to say that, since I started with ideal gases of temperature T, pressure p, and total volume V_f, and ended up with the same, then there was no entropy increase. On the other hand, the mixing is an irreversible process, and so there ought to be an entropy increase.

The standard textbook answer is: if the two gases are the same stuff — if they are not *distinguishable* — then no entropy increase has taken place. If, however, we start out with two different gases (the molecules of one might differ, say, in mass from those of the other, or perhaps one consists of positively charged ions, and the other of uncharged molecules), then we

end up with a mixture, and the total entropy increase is the entropy increase undergone by each gas separately.

$$\Delta S = n_1 R \log\left(\frac{V_f}{V_1}\right) + n_2 R \log\left(\frac{V_f}{V_2}\right). \tag{2.58}$$

This entropy increase is called the *entropy of mixing*.

Here's the argument that this is the right answer.

Let us take inspiration from the usual way of distinguishing mixtures from chemically pure substances, and say that the two gases are *separatable* if there is some quasistatic process that will unmix them and restore the system to its original state.[7] If , for example, they are molecules of different masses, then we might separate them in a centrifuge, and if they have different electric charge, we might separate them by putting them in an electric field. For simplicity, I will assume that there are differentially permeable membranes, that are impermeable to molecules of one gas, but not the other.[8]

If I fit a piston with one of these membranes, and couple the system to a heat bath of temperature T, and quasistatically push gas 1 back into its original volume, I do work on the system, which absorbs heat from the bath. Integrating $đQ/T$ along this process gives me an entropy change (a decrease) for the system equal to

$$\Delta S_1 = n_1 R \log\left(\frac{V_1}{V_f}\right). \tag{2.59}$$

I then push gas 2 back into its original volume with a piston that is impermeable to gas 2 but not to gas 1. The system undergoes an entropy change

$$\Delta S_2 = n_2 R \log\left(\frac{V_2}{V_f}\right). \tag{2.60}$$

I have therefore restored the original state via a quasistatic process, and hence I know the difference in entropy — the "entropy of unmixing" — which is just the negative of the entropy of mixing.[9]

[7] "Separable" already has too many meanings.

[8] As Daub (1969, p. 329) points out, the device of a membrane permeable to one gas but not the other, now a staple of textbook expositions, dates back to Boltzmann (1878).

[9] Of course, given the differentially permeable pistons, I can mimic the mixing process quasistatically, and get the same result. I did it this way to emphasize the key assumption — that there is some process that differentiates between the two gases.

What counts here is whether there is some process that acts differentially on the two gases, that could in principle used to unmix them. This clearly won't be the case if the gases consist of identical molecules. Hence having gases be unseparatable is a symmetry of the laws of physics — physical interactions treat the molecules of two unseparatable gases the same.

2.5 Helmholtz Free Energy

3 Kinetic Theory and Reduction

3.1 Principle Theories and Constructive Theories

In an article written in 1919 for *The London Times*, Einstein wrote,

> We can distinguish various kinds of theories in physics. Most of them are constructive. They attempt to build up a picture of the more complex phenomena out of the materials of a relatively simple formal scheme from which they start out. Thus the kinetic theory of gases seeks to reduce mechanical, thermal, and diffusional processes to movements of molecules—i.e., to build them up out of the hypothesis of molecular motion. When we say that we have succeeded in understanding a group of natural processes, we invariably mean that a constructive theory has been found which covers the processes in question.
>
> Along with this most important class of theories there exists a second, which I will call "principle-theories." These employ the analytic, not the synthetic, method. The elements which form their basis and starting-point are not hypothetically constructed but empirically discovered ones, general characteristics of natural processes, principles that give rise to mathematically formulated criteria which the separate processes or the theoretical representations of them have to satisfy. Thus the science of thermodynamics seeks by analytical means to deduce necessary conditions, which separate events have to satisfy, from the universally experienced fact that perpetual motion is impossible.
>
> The advantages of the constructive theory are completeness, adaptability, and clearness, those of the principle theory are logical perfection and security of the foundations. The theory of relativity belongs to the latter class. (Einstein, 1954, p. 223)

Note that Einstein says that we don't say that we understand something until we have a constructive theory. If we want to understand *why* materials obey the laws of thermodynamics, then, according to Einstein, this comes only with a constructive theory.

3.2 Elementary Kinetic Theory of Gases

The basic idea of the kinetic theory is simple: gases consist of a large number of discrete molecules, which interact only weakly except at short distances, at which they repel each other. Let us model such a gas by molecules that don't interact except via elastic collisions.

Suppose that we have such a gas, in a container which, for simplicity, we will take to have vertical sides and horizontal top and bottom. Let us ask what the pressure exerted by the gas on the underside of the lid of the container will be.

Let the container have volume V, lid of area A, and contain N molecules each of which has mass m. Let the position and velocity of the ith molecule be $(\mathbf{x}_i, \mathbf{v}_i)$, for $i = 1, ..., N$.

A particle that bounces off the lid will undergo a momentum change equal in magnitude to

$$\Delta P = 2mv_z. \tag{3.1}$$

We will estimate the pressure = force per unit area by considering a time interval δt and calculating

$$p = \frac{\sum_i \Delta P_i}{A\delta t}, \tag{3.2}$$

where the sum is taken over all the molecules that bounce off the lid during the time δt.

We will assume that the positions of the molecules are approximately evenly distributed, so that, for a subvolume V' of macroscopic size, the fraction of molecules in that subvolume is approximately

$$\frac{N'}{N} = \frac{V'}{V}. \tag{3.3}$$

(Note that this can't be exactly true for all subvolumes, and the approximation will tend to get worse as we consider smaller subvolumes.) Define $\rho = Nm/V$ as the average mass density of the gas. Then the fraction of all molecules that lie in a subvolume V' is approximately

$$\frac{N'}{N} = \left(\frac{\rho}{Nm}\right) V'. \tag{3.4}$$

Let $\Pi = \{\pi_k\}$ be a partition of the range of possible values of v_z into small intervals, and choose one particular element of this partition, $\pi_k = [v_z, v_z + \delta v_z]$. Let N_k be the number of molecules with z-component of velocity lying

in π_k. We will assume independence of the velocity and position distributions: that is, for this interval π_k (or any other interval we might have chosen), the number of molecules that have v_z in π_k *and* lie in a subvolume V' is

$$N_k \left(\frac{\rho}{Nm}\right) V'. \tag{3.5}$$

We ask: how many of these will collide with the lid, during our time interval δt? Take δt sufficiently small that we can disregard molecules that undergo other collisions during this time. Then a molecule will collide with the lid if and only if it is a distance less than $v_z \delta t$ from the lid. This picks out a region of volume $V_k = A v_z \delta t$. Let N_k^c be the number of molecules with $v_z \in \pi_k$ that are also in this collision region. The assumption of approximately uniform density, independent of velocity, gives us

$$\frac{N_k^c}{N_k} = \frac{V_k}{V} = \left(\frac{\rho}{Nm}\right) V_k. \tag{3.6}$$

Each of these undergoes momentum change $\Delta P = 2mv_z$, so the total momentum imparted by molecules with velocity in π_k is

$$(\Delta P)_k = N_k^c (2mv_z) = 2\rho (A\delta t) \left(\frac{N_k}{N}\right) v_z^2, \tag{3.7}$$

giving us a contribution to the pressure due to molecules in this velocity interval is equal to

$$\frac{(2mv_z)N_k^c}{A\delta t} = 2\rho \left(\frac{N_k}{N}\right) v_z^2. \tag{3.8}$$

To get the total pressure, we sum over the intervals in our partition with positive v_z (the others don't collide with the lid).

$$p = 2\rho \sum_{\pi_k \in \Pi, v_z > 0} \left(\frac{N_k}{N}\right) v_z^2. \tag{3.9}$$

We now assume that the distribution of v_z is symmetric under reflection: the number of molecules having velocity in the interval $[v_z, v_z + \delta v_z]$ is equal to the number with velocity in the interval $[-v_z - \delta v_z, -v_z]$. Then the sum in (3.9) will be just half the sum over the entire partition, and so

$$p = \rho \sum_{\pi_k \in \Pi} \left(\frac{N_k}{N}\right) v_z^2 = \rho \langle v_z^2 \rangle, \tag{3.10}$$

where $\langle v_z^2 \rangle$ denotes the mean value of v_z^2. We now assume isotropy of the velocity distribution,

$$\langle v_x^2 \rangle = \langle v_y^2 \rangle = \langle v_z^2 \rangle = \frac{1}{3} \langle v^2 \rangle. \tag{3.11}$$

This gives us the nice relation

$$p = \frac{1}{3} \rho \, \langle v^2 \rangle. \tag{3.12}$$

Recalling that $\rho = Nm/V$, we get

$$pV = \frac{1}{3} N \langle m v^2 \rangle = \frac{2}{3} N \langle K \rangle, \tag{3.13}$$

where $\langle K \rangle$ is the mean kinetic energy of the molecules. Comparison with the ideal gas law,

$$pV = nRT \tag{3.14}$$

gives us

$$pV = \frac{2}{3} N \langle K \rangle = nRT, \tag{3.15}$$

or,

$$T = \frac{2}{3} \left(\frac{N}{nR} \right) \langle K \rangle = \frac{N_A}{R} \langle K \rangle, \tag{3.16}$$

where N_A is Avogadro's number, the number of molecules per mole of gas. Defining $k = R/N_A$ (Boltzmann's constant), we get the result that the mean kinetic energy of the molecules is

$$\langle K \rangle = \frac{3}{2} \, k \, T. \tag{3.17}$$

We are thus led to construe the temperature of a gas as proportional to the mean kinetic energy of the molecules of the gas.

3.2.1 Heat capacity of a monatomic ideal gas

Suppose that we have a monatomic ideal gas—the molecules have no internal degrees of freedom, and so the total internal energy of the gas is just its total kinetic energy. That is,

$$U = N \langle K \rangle. \tag{3.18}$$

Comparison with (3.17) gives

$$U = \frac{3}{2}nRT = nc_vT, \tag{3.19}$$

which gives the result that the molar specific heat of a monatomic ideal gas is

$$c_V = \frac{3}{2}R. \tag{3.20}$$

This gives, in turn,

$$c_p = c_V + R = \frac{5}{2}R, \tag{3.21}$$

and

$$\gamma = \frac{c_p}{c_V} = \frac{5}{3}. \tag{3.22}$$

4 The Second Law Revised

4.1 Tensions between thermodynamics and kinetic theory

In the previous section, we saw some promising first steps towards the reduction of the thermodynamics of gases to kinetic theory. However, here are some tensions between the kinetic theory of gases and thermodynamics as we have been conceiving it.

- We have been treating of equilibrium as a state in which nothing happens (*cf.* quote from Atkins at beginning of §2.-1). On the kinetic theory, an equilibrium state is a state that is seething with activity.

- We have been treating a gas in equilibrium as if it is a uniform substance, with uniform temperature and pressure throughout. On a microscopic level, however, it is far from uniform! The pressure exerted on the sides of its container is neither steady nor uniform — though it will *average out* to something approximately steady and uniform, as long as we consider areas large enough and times long enough that a large number of molecular collisions are involved.

- We have taken the distinction between energy transfer as work, and energy transfer as heat, to be a clear one. On the kinetic theory, however, to heat something is to convey kinetic energy to its molecules. The difference becomes: when I do work on a system, say, by moving a piston, the parts of the piston move in an orderly fashion, all in the same direction, whereas, when I heat something, the added motion of the molecules is scattered in a higgedly-piggledy fashion. Is this a distinction that holds up at the molecular level?

- Thermodynamics distinguishes between reversible and irreversible processes. The molecular dynamics are TRI (unless those need revision, too).

The last is the most serious, and is what convinced people that thermodynamics, as conceived, wasn't quite right, and what was to be reduced was a revised version.

4.2 The Reversibility Argument

In the decade 1867-1877, the major figures working on the kinetic theory came to realize that the 2nd law of thermodynamics, as it had been conceived, was not to be recovered from the kinetic theory. At best one could recover a weaker version that could nonetheless account for the empirical evidence in favour of the 2nd law as originally conceived. Moreover, it seemed that some notion of probability was required; what the original version deemed impossible was to be regarded as merely highly improbable.

It was considerations of reversibility of molecular dynamics that led to these conclusions. If the molecular dynamics are TRI, then the temporal inverse of any dynamically possible process is also dynamically possible, including those that are regarded as thermodynamically irreversible.

On a letter from Maxwell dated Dec. 11, 1867 (the letter in which Maxwell introduced Tait to the creature came to be called "Maxwell's demon"[10]), P. G. Tait wrote, "Very good. Another way is to reverse the motion of every particle of the Universe and to preside over the unstable motion thus produced" (Knott, 1911, p. 214).

The reversibility argument is spelled out in a letter, dated Dec. 6, 1870, from Maxwell to John William Strutt, Baron Rayleigh; Maxwell follows this with an exposition of the demon, and then draws the

> *Moral.* The 2nd law of thermodynamics has the same degree of truth as the statement that if you throw a tumblerful of water into the sea, you cannot get the same tumblerful of water out again (Garber et al., 1995, p. 205).

Maxwell's view is that processes that, from the point of view of thermodynamics, are regarded as irreversible, are ones whose temporal inverses are not impossible, but merely improbable. In a letter to the editor of the *Saturday Review*, dated April 13, 1868, Maxwell draws an analogy between mixing of gases and balls shaken in a box.

[10] Thomson attributes the name to Maxwell:

> The definition of a "demon", according to the use of this word by Maxwell, is an intelligent being endowed with free will, and fine enough tactile and perceptive organisation to give him the faculty of observing and influencing individual molecules of matter (Thomson, 1874, p. 441).

But Maxwell says that it was Thomson who gave the creatures this name (Knott, 1911, p. 214).

> As a simple instance of an irreversible operation which (I think) depends on the same principle, suppose so many black balls put at the bottom of a box and so many white above them. Then let them be jumbled together. If there is no physical difference between the white and black calls, it is exceedingly improbable that any amount of shaking will bring all the black balls to the bottom and all the white to the top again, so that the operation of mixing is irreversible unless either the black balls are heavier than the white or a person who knows white from black picks them and sorts them.
>
> Thus if you put a drop of water into a vessel of water no chemist can take out that identical drop again, though he could take out a drop of any other liquid (in Garber et al. 1995, 192–193).

We find similar considerations in Gibbs several years later,

> when such gases have been mixed, there is no more impossibility of the separation of the two kinds of molecules in virtue of their ordinary motions in the gaseous mass without any external influence, than there is of the separation of a homogeneous gas into the same two parts into which it as once been divided, after these have once been mixed. In other words, the impossibility of an uncompensated decrease of entropy seems to be reduced to improbability (Gibbs 1875, 229; 1961, 167).

It was Loschmidt who, in 1876, drew Boltzmann's attention to reversibility considerations. In his response to Loschmidt, Boltzmann (1877) acknowledged that there could be no purely dynamical proof of the increase of entropy.[11]

It is one thing to acknowledge that violations of the second law will sometimes occur, albeit with low probability. Maxwell went further, asserting that, on the small scale, minute violations of the second law will continually occur; it is only large-scale, observable violations that are improbable.

> the second law of thermodynamics is continually being violated, and that to a considerable extent, in any sufficiently small group of molecules belonging to a real body. As the number of molecules

[11]For further discussion of the probabilistic turn in Boltzmann's thinking, see Uffink (2007b), Brown et al. (2009).

> in the group is increased, the deviations from the mean of the whole become smaller and less frequent; and when the number is increased till the group includes a sensible portion of the body, the probability of a measurable variation from the mean occurring in a finite number of years becomes so small that it may be regarded as practically an impossibility.
>
> This calculation belongs of course to molecular theory and not to pure thermodynamics, but it shows that we have reason for believing the truth of the second law to be of the nature of a strong probability, which, though it falls short of certainty by less than any assignable quantity, is not an absolute certainty (Maxwell 1878b, p. 280; Niven 1965, pp. 670–71).

What is accepted by most physicists today, and goes by the name of the 2nd law of thermodynamics, is something along the lines of

> Although fluctuations will occasionally result in heat passing spontaneously from a colder body to a warmer body, these fluctuations are inherently unpredictable and it is impossible for there to be a process that *consistently and reliably* harnesses these fluctuations to do work.

Call this the *probabilistic version of the second law of thermodynamics*.

4.3 The Maxwellian View

4.3.1 A Third Second Law

Maxwell placed a further limitation on the 2nd law. For Maxwell, even the probabilistic version of the 2nd law holds only so long as we are in a situation in which molecules are dealt with only *en masse*. This is the limitation of which he speaks, in the section of *Theory of Heat* that introduces the demon to the world.

> One of the best established facts in thermodynamics is that it is impossible in a system enclosed in an envelope which permits neither change of volume nor passage of heat, and in which both the temperature and the pressure are everywhere the same, to produce any inequality of temperature or pressure without the

> expenditure of work. This is the second law of thermodynamics, and it is undoubtedly true as long as we can deal with bodies only in mass, and have no power of perceiving the separate molecules of which they are made up. But if we conceive of a being whose faculties are so sharpened that he can follow every molecule in its course, such a being, whose attributes are still as essentially as finite as our own, would be able to do what is at present impossible to us. For we have seen that the molecules in a vessel full of air at uniform temperature are moving with velocities by no means uniform, though the mean velocity of any great number of them, arbitrarily selected, is almost exactly uniform. Now let us suppose that such a vessel is divided into two portions, A and B, by a division in which there is a small hole, and that a being, who can see the individual molecules, opens and closes this hole, so as to allow only the swifter molecules to pass from A to B, and only the slower ones to pass from B to A. He will thus, without expenditure of work, raise the temperature of B and lower that of A, in contradiction to the second law of thermodynamics.
>
> This is only one of the instances in which conclusions which we have drawn from our experience of bodies consisting of an immense number of molecules may be found not to be applicable to the more delicate observations and experiments which we may suppose made by one who can perceive and handle the individual molecules which we deal with only in large masses. (Maxwell, 1871, pp. 308–309).

Note that there is in this no hint that there might be some principle of physics that precludes the manipulations of the demon, or constrains it to dissipate sufficient energy that the net change of entropy it produces is positive. Moreover, Maxwell leaves it open that the requisite manipulations might become technologically possible in the future—the demon does what is *at present* impossible for us. What Maxwell is proposing, as a successor to the second law, is strictly weaker than the probabilistic version. For Maxwell, even the probabilistic version is limited in its scope—it holds only in circumstances in which there is no manipulation of molecules individually or in small numbers.

4.3.2 Maxwell on thermodynamics

Maxwell's conception of the status of the second law ties in with his conception of the status and purpose of the science of thermodynamics.

Central to thermodynamics is a distinction between two ways in which energy can be transferred from one system to another: it can be transferred as heat, or else one system can do work on the other. The second law of thermodynamics requires, for its very formulation, a distinction between these two modes of energy transfer. In Clausius' formulation,

> Heat cannot pass from a colder body to a warmer body without some other change connected with it occurring at the same time.[12]

To see that this hangs on a distinction between heat and work, note that it becomes false if we don't specify that the energy is transferred as heat. It is not true that *no* energy can be conveyed from a cooler body to a warmer body without some other change connected with it: if two gases are separated by an insulating movable piston, the gas at higher pressure can compress— that is, do work on— the gas at lower pressure, whatever their respective temperatures.

The Kelvin formulation of the second law is,

> It is impossible, by means of inanimate material agency, to derive mechanical effect from any portion of matter by cooling it below the temperature of the coldest of the surrounding objects (quoted in Uffink 2001, p. 327).

This statement does *not* say that we cannot cool a body below the temperature of the coldest surrounding objects. Refrigerators are possible. The difference is: though we can derive mechanical effect—that is, do work— by extracting heat from a hotter body, using some of the energy to do work, and discarding the rest into a cooler reservoir, extraction of heat from a body that is already cooler than any body that might be used as a reservoir requires the opposite of deriving mechanical effect: it requires us to use up some energy that could have been used for mechanical effect, in order to effect the transfer. Thus the Kelvin statement, also, requires a distinction between deriving mechanical effect from a body and extracting heat from it.

[12] "Es kann nie Wärme aus einem kälteren Körper übergehen, wenn nicht gleichzeitig eine andere damit zusammenhängende Aenderung eintritt." Quoted by Uffink (2001, p. 333).

What is this distinction? On the kinetic theory of heat, when a body is heated, the total kinetic energy of its molecules is increased, so, for body A to heat body B, parts of A must interact with parts of B to change their state of motion. When A does work on B, it is again the case that parts of A act on parts of B to change their state of motion. The difference is: in heat transfer, energy is transferred to the parts of the body in a haphazard way; the resulting motions cannot be tracked. This limits our ability to recover the energy as work.

Put this way, the distinction seems to rest on anthropocentric considerations, or, better, on consideration of the means we have available to us for keeping track of and manipulating molecules. We shall call considerations that turn on the means available to an agent for gathering information about a system or for manipulating it *means-relative*; these are matters that can vary between agents, but it would be misleading to call them subjective, as we are considering limitations on the physical means that are at the agents' disposal. On Maxwell's view, the distinction between work and heat is means-relative.

> Available energy is energy which we can direct into any desired channel. Dissipated energy is energy we cannot lay hold of and direct at pleasure, such as the energy of the confused agitation of molecules which we call heat. Now, confusion, like the correlative term order, is not a property of material things in themselves, but only in relation to the mind which perceives them. A memorandum-book does not, provided it is neatly written, appear confused to an illiterate person, or to the owner who understands thoroughly, but to any other person able to read it appears to be inextricably confused. Similarly the notion of dissipated energy could not occur to a being who could not turn any of the energies of nature to his own account, or to one who could trace the motion of every molecule and seize it at the right moment. It is only to a being in the intermediate stage, who can lay hold of some forms of energy while others elude his grasp, that energy appears to be passing inevitably from the available to the dissipated state (Maxwell 1878a, p. 221; Niven 1965, p. 646).

That there is some energy that, for us, counts as dissipated energy has to do, according to Maxwell, with the large number and small size of the molecules that make up a macroscopic body.

> The second law relates to that kind of communication of energy which we call the transfer of heat as distinguished from another kind of communication of energy which we call work. According to the molecular theory the only difference between these two kinds of communication of energy is that the motions and displacements which are concerned in the communication of heat are those of molecules, and are so numerous, so small individually, and so irregular in their distribution, that they quite escape all our methods of observation; whereas when the motions and displacements are those of visible bodies consisting of great numbers of molecules moving all together, the communication of energy is called work (Maxwell 1878b, p. 279; Niven 1965, p. 669).

If heat and work are means-relative concepts, then perforce so is entropy. The entropy difference between two equilibrium states of a system is given by

$$\Delta S = \int \frac{đQ}{T},$$

where the integral is taken over any quasistatic process joining the two states, and $đQ$ is the increment in heat absorbed from the system's environment. Thus, on Maxwell's view, the very concepts required to state the second law of thermodynamics are means-relative. For more on the Maxwellian view of thermodynamics and statistical mechanics, see Myrvold (2011).

4.4 Exorcising Maxwell's Demon

If a machine is possible that could behave as a Maxwell Demon, and if this machine could operate without a compensating increase of entropy either in its own internal state or in some auxiliary system, then even the probabilistic version of the Second Law is false. Most physicists believe that such a machine is, indeed, impossible. There is less consensus on why.

To operate reliably, the Demon must:

1. Make an observation regarding the position and velocity of an incoming molecule,

2. Record the result, at least temporarily,

3. Operate the trap-door dividing the two sides of the vessel of gas,

4. If operating reversibly, erase the record of the observation.

If the statistical version of the Second Law is not to be violated, then this series of activities must, on average, generate an increase of entropy at least as great as the entropy decrease effected in the gas. In 1929, Szilard, basing his calculation on a simple one-molecule engine, concluded,

> If an information processing system gains information sufficient to discriminate among n equally likely alternatives, this must on average be accompanied by an entropy increase of at least $k \log n$.

This is known as *Szilard's Principle*. In 1961, Landauer concluded

> If an information processing system erases information sufficient to discriminate among n equally likely alternatives, this must on average be accompanied by an entropy increase of at least $k \log n$.

This is known as *Landauer's Principle*.

These principles are argued for *on the assumption of the correctness of the statistical version of the Second Law*. They may be correct, but invoking them would not help convince a modern Maxwell, skeptical of the Second Law, of its correctness. For more discussion of Maxwell Demons, see Earman and Norton (1998, 1999); Leff and Rex (2003); Norton (2005, 2011).

5 The Boltzmann H-theorem and its discontents

A simple kinetic model of the ideal gas has led to the identification of thermodynamic variables with aggregate properties of molecules: pressure as the average force per unit area exerted on the sides of the container (or object suspended in the gas), temperature as proportional to mean kinetic energy of the molecules. We want more: we want construal of entropy in molecular terms, and an explanation of the approach to equilibrium.

Boltzmann's H-theorem was an important step towards the latter. Boltzmann considered how the distribution of the velocities of the molecules of a gas could be expected to change under collisions, argued that there was a unique distribution—now called the Maxwell-Boltzmann distribution—that was stable under collisions, and, moreover, that a gas that initially had a different distribution

Maxwell-Boltzmann distribution: the fraction of molecules having velocity in a small volume in $\mathbf{v}$-space, $[v_x, v_x + \delta v_x] \times [v_y, v_y + \delta v_y] \times [v_z, v_z + \delta v_z]$, is proportional to

$$e^{-mv^2/2kT}\, \delta v_x \delta v_y \delta v_z. \tag{5.1}$$

To argue that this was the unique equilibrium, which would be approached if we started with a different distribution, Boltzmann defined a quantity, (now) called H, showed that it reached a minimum value for this distribution, and argued that it would decrease to its minimum.[13]

5.1 The Ehrenfest wind-tree model

Rather than go into the intricacies of the H-theorem, we will look at a simple toy model (the 'wind-tree model') due to the Ehrenfests (Ehrenfest and Ehrenfest, 1990, pp. 10–13) , that is easy to analyze but share some salient features with Boltzmann's gases.

The model is two-dimensional. Molecules of 'wind' move in a plane. All have same speed v, and each moves in one of four directions: north, south, east, west. Scattered randomly throughout the plane are square obstacles—'trees'— which don't move, with sides of length a. These have their diagonals

[13] Why 'H'? The quantity (as we shall see) is related to the entropy, and indeed, Boltzmann originally used 'E'. There is some reason to believe that H is meant to be a capital η. See Hjalmars (1977), if you care.

aligned with the north-south and east-west directions, so that, when a wind molecule hits one, it is deflected into one of the other four directions. The trees have uniform mean density n per unit volume.

We want to know how a given initial distribution of wind-velocities will change with time. Consider a small time δt. A given wind-molecule, say, one moving east, will change velocity during this time iff it hits a tree, in which case it will be deflected either to the north or to the south. For any given tree, the region containing wind-molecules which will be deflected to the north has area $a\, v\delta t$. We can therefore talk about east-north collision areas, *etc.*

We assume:

> The fraction of all east-travelling molecules that happen to lie in east-north collision strips is equal to the proportion of the plane occupied by such strips.

And, of course, we make the corresponding assumption for all the other directions. Call this the *Stoßzahlansatz*: "collision-number assumption."[14] Note that we made a similar assumption regarding velocity distribution in our derivation of the pressure of an ideal gas.

The proportional area of collision-strips of each type is $nav\,\delta t$, where n is the number of trees per unit area. Define $\alpha = nav$. Let $f_i = n_i/N$ (where i takes as values n, s, e, w) be the fraction of all wind-molecules travelling in the i-direction. From the *Stoßzahlansatz* it follows that, in a time δt, a fraction $2\alpha\,\delta t$ of east-moving molecules will be deflected into other directions. At the same time a fraction $\alpha\,\delta t$ of north-moving molecules will be deflected into the east direction, and the same fraction of south-moving molecules. This gives us

$$\delta f_e = (-2\alpha f_e + \alpha f_n + \alpha f_s)\,\delta t. \tag{5.2}$$

This, and corresponding considerations for the other directions, gives us the system of equations,

$$\begin{aligned} df_e/dt &= -2\alpha f_e + \alpha(f_n + f_s) \\ df_w/dt &= -2\alpha f_w + \alpha(f_n + f_s) \\ df_n/dt &= -2\alpha f_n + \alpha(f_e + f_w) \\ df_s/dt &= -2\alpha f_s + \alpha(f_e + f_w) \end{aligned} \tag{5.3}$$

[14] The translator of the Ehrenfests' book left certain key terms—*Stoßzahlansatz*, *Umkehreinwand*, *Wiederkehreinwand*—untranslated, and the tradition among commentators has been to follow suit.

Inspection of the system of equations (5.3) shows that a stationary solution is obtained when all the f_i's are equal; that is,

$$f_e = f_n = f_w = f_s = \frac{1}{4}. \tag{5.4}$$

It's not hard to get an explicit solution for our system of equations, given arbitrary initial frequencies (details in §5.1.2, below). The upshot is: whatever the initial frequencies are, we get an exponential approach to the equilibrium state.

5.1.1 The wind-tree H-theorem

Define

$$H = N \sum_i f_i \log f_i, \tag{5.5}$$

where the sum is taken over the directions for which $f_i \neq 0$.[15]

Because, for all i, $0 \leq f_i \leq 1$, H cannot be positive. It takes its maximum value of 0 when all the molecules are going the same direction; that is, one of the f_i's is 1, and all the others 0. It takes its minimum value in the equilibrium state, when all the f_i's are the same. H is, therefore, in some sense an estimate of how close the gas is to equilibrium.

Consider the rate of change of this quantity,

$$\begin{aligned} dH/dt &= N \sum_i (df_i/dt) \log f_i + N \sum_i df_i/dt \\ &= N \sum_i (df_i/dt) \log f_i. \end{aligned} \tag{5.6}$$

A bit of algebra gives us

$$\begin{aligned} dH/dt = -N\alpha \Bigg[&(f_e - f_n) \log \left(\frac{f_e}{f_n} \right) + (f_e - f_s) \log \left(\frac{f_e}{f_s} \right) \\ &+ (f_w - f_n) \log \left(\frac{f_w}{f_n} \right) + (f_w - f_s) \log \left(\frac{f_w}{f_s} \right) \Bigg] \end{aligned} \tag{5.7}$$

The quantity in square brackets in (5.7) is non-negative, because, for any positive x, y,

$$(x - y) \log \left(\frac{x}{y} \right) \geq 0. \tag{5.8}$$

[15] If you like, you can think of this as adopting the convention: $0 \times \log 0 = 0$.

Thus, we can conclude, that, in any state,

$$\frac{dH}{dt} \leq 0, \tag{5.9}$$

Moreover, $dH/dt = 0$ only in the equilibrium state $f_i = const.$ This gives us a simple way to conclude that, for arbitrary initial frequencies, the state approaches the equilibrium state, where it takes its minimum value.

This is a temporally asymmetric conclusion. The evolution of the frequencies expressed by equations (5.3) is not invariant under time-reversal. But the underlying dynamics of collisions *is* TRI. This means that we must have used an assumption to get from the dynamics to the equations (5.3) that introduces a temporal asymmetry. But the only assumption we used was the *Stoßzahlansatz* . And, indeed, it is a temporally asymmetric assumption (if we're not in the equilibrium state).

This is easiest to see if we consider the extreme disequilibrium state. Suppose that, at $t = 0$, $f_e = 1$, and that the *Stoßzahlansatz* holds. A time δt later, a fraction $2\alpha\delta t$ of the molecules have been scattered, half into the north-direction, half into the south. Suppose, now, we reverse the velocities, and ask what fraction of, say, the north-travelling molecules will collide with a tree in time δt. Answer: all of them! The wind-molecules that are not travelling in the west direction are all on collision courses that will turn them into west-travelling molecules.

We may find the *Stoßzahlansatz* to be a reasonable assumption in the forward direction, but its temporal reverse is certainly *not*, unless the state is already an equilibrium state. The intuition here seems to have something to do with causality. We might expect the velocity of a molecule to be independent of what interactions lie to its near future. We *don't* expect its velocity to be independent of interactions in its recent past. This suggests an attitude towards the temporal asymmetry of thermodynamics that grounds thermodynamic asymmetry on the temporal asymmetry of causality. This seems a promising avenue, though it is not one that has been popular; see, however, Oliver Penrose (2001).[16]

[16]Don't confuse Oliver Penrose, who specializes in statistical mechanics, with his brother Roger, who also has views on temporal asymmetry!

5.1.2 Explicit solution for wind-tree model

It's handy to write the system of equations (5.3) in matrix form. Writing $\mathbf{f} = (f_e, f_w, f_n, f_s)$, let us rewrite our equations as,

$$d\mathbf{f}/dt = C\mathbf{f}. \tag{5.10}$$

where

$$C = \begin{pmatrix} -2\alpha & 0 & \alpha & \alpha \\ 0 & -2\alpha & \alpha & \alpha \\ \alpha & \alpha & -2\alpha & 0 \\ \alpha & \alpha & 0 & -2\alpha \end{pmatrix} \tag{5.11}$$

The solution is give by

$$\mathbf{f}(t) = U(t)\,\mathbf{f}_0, \tag{5.12}$$

Comparison of (5.10) and (5.12) tells us that

$$dU/dt = CU. \tag{5.13}$$

You can easily verify that this is solved by

$$U(t) = \frac{1}{4}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} + \frac{1}{4}\begin{pmatrix} 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & 1 \end{pmatrix} e^{-4\alpha t} + \frac{1}{2}\begin{pmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & 1 \end{pmatrix} e^{-2\alpha t} \tag{5.14}$$

As t increases, this decays exponentially to

$$U_\infty = \frac{1}{4}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}, \tag{5.15}$$

which takes arbitrary initial frequencies into the equilibrium state.

5.2 Boltzmann's H-theorem

In his original presentation (1872) of the H-theorem, Boltzmann gave the impression that the approach to equilibrium was a consequence of the laws of mechanics alone, applied to molecular collisions. This can't be right! Two types of objection showed that it can't be right:

- **The reversibility objection** (*Umkehreinwand*).[17] For any set of trajectories of the molecules of a gas, the time-reversed trajectories are also possible. Hence, a monotonic decrease of *anything* can't be a consequence of the microphysical dynamics alone.

- **The recurrence objection** (*Wiederkehreinwand*). (Raised by Zermelo.) Take a classical system, confined to a bounded phase space (*e.g.* a gas in a box, with fixed total energy). Consider a small open neighbourhood of the initial state, and ask, will the state of the system, after it leaves that neighbourhood, ever return to it? Poincaré's recurrence theorem answers: yes, for almost all initial phase space-points, where 'almost all' means 'all except for a set of Lebesgue measure zero.[18] This has the paradoxical-sounding consequence that a gas that starts out in a non-equilibrium state, though it may relax to equilibrium in a short time, will, if left to itself long enough, almost certainly return arbitrarily close to its initial state.

Can this possibly true? Would a box of gas, left on its own, eventually spontaneously end up occupying only half of the box? Lest this sound like mathematics gone mad, recall that we have already acknowledged, as an unsurprising consequence of the kinetic theory, the existence of small pressure fluctuations. The Poincaré recurrence theorem says that, if you wait long enough, all sorts of pressure fluctuations occur in an isolated gas combined to a box. In his book on non-eq SM, Dorfman provides some illuminating calculations of Poincaré recurrence times. Following is a table of the time you would have to wait before you would expect to see a 1% fluctuation of the pressure in a sphere of radius a of a gas at standard temperature and pressure(Dorfman, 1999).

a (cm.)	t_r (sec.)
10^{-5}	10^{-11}
2.5×10^{-5}	1
3×10^{-5}	10^{6}
5×10^{-5}	10^{68}
1	$10^{10^{14}}$

[17] Credited by the Ehrenfests to Loschmidt, though there had been some discussion of this among the English physicists.

[18] If the phrase 'Lebesgue measure' is unfamiliar, your intuitive idea of phase space volume will suffice.

The reversibility objection applies to our wind-tree model. The recurrence objection does not, in the original formulation, which has the wind-tree forest spread out over an infinite plane. It would apply if we put it in a finite box.

What do these two objections show? First of all, they emphasize that the monotonic decrease of H to its equilibrium value is not a consequence of the molecular dynamics alone. Something else is needed: the *Stoßzahlansatz*. Moreover, they tell us something about this assumption. Away from equilibrium, it cannot hold in both temporal directions. And, even it holds at some initial time, it cannot remain true indefinitely, for a bounded, isolated system.

Intuitively: even if the *Stoßzahlansatz* is valid at some particular time t_0,

- Though the state of a molecule at t_0 is independent of what lies ahead of it in its near future (and will continue to be at least approximately so, for some time), its state shortly after t_0 will not be independent of interactions in its recent past.

- Over long enough time periods, as molecules encounter molecules with which they have interacted, the correlations built up via past interactions become relevant, and so the *Stoßzahlansatz* cannot be regarded as a reasonable approximation arbitrarily far into the future.

5.3 The significance of H

5.3.1 Wind-tree H

The detailed microscopic state of the wind-tree gas requires specification of position and velocity of each of the individual molecules. We have represented the macro-state by specification of the numbers $\{n_e, n_w, n_n, n_s\}$ of molecules with each v-direction. If there are N wind-molecules in total, for a given macrostate $\{n_e, n_w, n_n, n_s\}$ the number of distinct distributions of wind-molecules among the four directions that agree with this macrostate is

$$\Omega = \frac{N!}{n_e!\, n_w!\, n_n!\, n_s!}. \tag{5.16}$$

This gives

$$\log \Omega = \log N! - \sum_i \log n_i! \tag{5.17}$$

To get a handle on this, apply Stirling's approximation formula,

$$\log N! \approx N \log N - N. \tag{5.18}$$

This gives us (remembering that $\sum_i n_i = N$),

$$\begin{aligned} \log \Omega &\approx N \log N - \sum_i n_i \log n_i \\ &= N \left(\log N - \sum_i f_i \log(N f_i) \right) \\ &= -N \sum_i f_i \log f_i = -H, \end{aligned} \tag{5.19}$$

or,

$$H \approx -\log \Omega. \tag{5.20}$$

5.3.2 Boltzmann's H and Boltzmann's entropy

Consider a gas made up of N molecules, each having r degrees of freedom. The phase space of each molecule, called μ-space, is $2r$-dimensional. The phase space of the gas, called Γ-space, is $2Nr$-dimensional.

Let us partition μ-space into compartments $\{\omega_i\}$, of equal volume $[\omega]$. For any state of the gas (that is, any point $x \in \Gamma$) let n_i be the number of molecules that, when the gas is in state x, occupy ω_i. Consider the set (called a 'Z-star') of points in Γ that have the same occupation-numbers as x. The number of distinct ways to achieve a given $\mathbf{n} = \{n_i\}$ will be

$$\Omega = \frac{N!}{n_1!\, n_2! \ldots n_k!}, \tag{5.21}$$

where, of course, we only have to include occupation-numbers $\{n_i\}$ that happen to be nonzero. Therefore, the volume of the region of Γ-space sharing x's set of occupation numbers is

$$\Omega\, [\omega]^N = \frac{N!}{n_1!\, n_2! \ldots n_k!}\, [\omega]^N. \tag{5.22}$$

As before, we will have (for N sufficiently large for application of Stirling's approximation),

$$\log \Omega \approx -N \sum_i f_i \log f_i, \tag{5.23}$$

where $f_i = n_i/N$. Let $\mathcal{E}_i$ be the energy possessed by a molecule in μ-space compartment ω_i (we take these small enough that the energy can be regarded as roughly constant in ω_i). The total energy of the gas is therefore,

$$U = \sum_i n_i \, \mathcal{E}_i = N \sum_i f_i \, \mathcal{E}_i. \tag{5.24}$$

Suppose, now, we ask: subject to constraint (5.24), what will be the values of $\{f_i\}$ that maximize Ω? If we take N large, so that the f_i's can be treated as continuously varying quantities, then minimizing H (and hence maximizing Ω) is achieved by:

1. All compartments with the same energy $\mathcal{E}_k$ have the same occupation-number n_k.

2. $n_k \propto e^{-\beta \mathcal{E}_k}$, for some constant β.

Apply these ideas to an ideal gas—the energy of a molecule is kinetic energy only, hence independent of position. The set of occupation numbers that maximizes Ω will be those in which the molecules are evenly spread over the accessible volume. Also, we will have

$$\frac{1}{2} m \langle v^2 \rangle = \frac{3}{2} \, \beta^{-1}, \tag{5.25}$$

Recall that we decided (§3.2), that, for an ideal gas,

$$\langle K \rangle = \frac{3}{2} \, k \, T. \tag{5.26}$$

Thus, we identify the parameter β:

$$\beta = \frac{1}{kT}. \tag{5.27}$$

Therefore. the distribution of velocities is the Maxwell-Boltzmann distribution.

If we calculate H for such a maximum-Ω state, we get

$$H = -N \left(\frac{3}{2} \log T + \log V + const. \right) \tag{5.28}$$

Compare this with the entropy of a monatomic ideal gas,

$$\begin{aligned} S &= C_V \log T + nR \log V + const. \\ &= kN \left(\frac{3}{2} \log T + \log V + const. \right). \end{aligned} \tag{5.29}$$

So, we have, for a monatomic ideal gas,

$$S = -kH + const. \tag{5.30}$$

Define the *Boltzmann entropy*

$$S_B = k \log \Omega. \tag{5.31}$$

This has the properties:

- For an isolated gas (fixed total energy), if we ask for the distribution of molecular positions and velocities that maximizes Ω—that is, the one that takes up the largest volume of Γ-space—it's the one in which molecular positions are distributed evenly over the box, and velocities according to the Maxwell-Boltzmann distribution.
- In such a state, the functional dependance of S_B on temperature and volume is the same as that of the thermodynamic entropy S.

Without going into details of molecular interactions, it's at least plausible that molecular collisions will lead the system into 'typical' regions of its phase space—those that maximize S_B. There's a temptation to put aside molecular dynamics and argue on the basis solely of phase-space volume that systems will tend to wander into regions of higher Boltzmann entropy. But this can't be right. What we *can* hope for, however, is that such a conclusion will not depend too sensitively on the details of the molecular dynamics. More on this later.

5.3.3 Boltzmann entropy of an ideal gas

I exhibited the result (5.28) without the additive constants, because I wanted to emphasize the dependence on volume and temperature, and compare it with the thermodynamic entropy, which is defined only up to an additive constant. Given a choice of the size of the elements partition of phase space,

$[\omega]$, the formula (5.31) uniquely determines the additive constant. Let's do the detailed calculation.

Because it's easier (in this case, at least, because we know how to do Gaussian integrals) to do integrals than sums, we will approximate the distribution $\mathbf{f} = (f_i)$ by a continuous function f on μ-space that varies little within each partition-element ω_i.

$$f_i \approx f(\mathbf{x}_i, \mathbf{p}_i)\,[\omega] \tag{5.32}$$

where $(\mathbf{x}_i, \mathbf{p}_i)$ is a representative point in ω_i (since f varies little in this region, it won't matter much *which* points we take). Then

$$H = N \sum_i f_i \log f_i = N \sum_i [f(\mathbf{x}_i, \mathbf{p}_i) \log f(\mathbf{x}_i, \mathbf{p}_i)]\,[\omega] + N \log[\omega]. \tag{5.33}$$

Since we're interested in the values of f_i that maximize Ω, and the energy associated with point $(\mathbf{x}, \mathbf{p})$ in μ-space is $p^2/2m$, take

$$f(\mathbf{x}, \mathbf{p}) = C^{-1} e^{-\beta p^2/2m}, \tag{5.34}$$

where C is a normalization constant. The requirement that the integral of f over the spatial volume V, and all possible momenta, be equal to unity gives

$$C = \left(\frac{2\pi m}{\beta}\right)^{\frac{3}{2}} V. \tag{5.35}$$

We replace the sum in (5.33) with an integral,

$$\sum_i [f(\mathbf{x}_i, \mathbf{p}_i) \log f(\mathbf{x}_i, \mathbf{p}_i)]\,[\omega] \approx \int\int f(\mathbf{x}, \mathbf{p}) \log f(\mathbf{x}, \mathbf{p})\, d^3x\, d^3p = \langle \log f \rangle, \tag{5.36}$$

where $\langle\ \rangle$ denotes expectation value with respect to the measure f. We have,

$$\langle \log f \rangle = -\log C - \frac{\beta}{2m} \langle p^2 \rangle. \tag{5.37}$$

For this distribution,

$$\langle p^2 \rangle = \frac{3m}{\beta}, \tag{5.38}$$

or,

$$\langle K \rangle = \frac{\langle p^2 \rangle}{2m} = \frac{3}{2}\,\beta^{-1}, \tag{5.39}$$

on which we based our conclusion that $\beta = 1/kT$. So, we have

$$\langle \log f \rangle = -\left(\log C + \frac{3}{2} \right), \tag{5.40}$$

$$\begin{aligned} H &= -N \left(\log C + \frac{3}{2} - \log[\omega] \right) \\ &= -N \left[\log \left(\frac{V(mkT)^{3/2}}{[\omega]} \right) + \frac{3}{2} \log 2\pi + \frac{3}{2} \right], \end{aligned} \tag{5.41}$$

or

$$\begin{aligned} S_B &= k \log \Omega \approx -kH \\ &= kN \left[\log \left(\frac{V(mkT)^{3/2}}{[\omega]} \right) + \frac{3}{2} \log 2\pi + \frac{3}{2} \right]. \end{aligned} \tag{5.42}$$

5.3.4 Gibbs' paradox

The expression (5.42) for the Boltzmann entropy of a classical, monatomic ideal gas is what is yielded by a straightforward calculation. But there's something wrong with it—it's not an extensive quantity.

To see this, suppose that we have two samples of the same ideal gas at the same temperature and pressure, with N_1 and N_2 molecules, respectively, occupying volumes V_1 and V_2. The combined system has $N = N_1 + N_2$ molecules and occupies a volume $V = V_1 + V_2$. Applying (5.42) to these gases, we find,

$$S \neq S_1 + S_2. \tag{5.43}$$

In fact, we get (after a bit of algebra),

$$S - (S_1 + S_2) = kN_1 \log (V/V_1) + kN_2 \log (V/V_2). \tag{5.44}$$

So, if we initially have N_1 molecules of a gas, occupying V_1, and N_2 molecules, occupying V_2, and remove the partition between them, allowing them to mix, then applying (5.42) to calculate the entropy of the mixture, yields the result that this entropy exceeds the sum of the initial entropies by an amount equal to the entropy of mixing of two distinct gases (recall discussion in §2.4.1). This result—that a straightforward classical calculation results in an entropy of mixing, even for samples of the same gas—is called *Gibbs' paradox*.

The expression for the entropy of an ideal gas that one finds in textbooks is usually derived from a treatment of the gas as a set of quantum particles confined to a box. This results in the *Sackur-Tetrode formula*,

$$S_{ST} = kN\left[\log\left(\frac{V(mkT)^{3/2}}{Nh^3}\right) + \frac{3}{2}\log 2\pi + \frac{5}{2}\right]. \tag{5.45}$$

The N in the denominator renders it extensive: if I have samples of gas at the same temperature and pressure, so that V_i/N_i is the same for all i, then

$$S_{ST}(N, V, T) = \sum_i S_{ST}(N_i, V_i, T). \tag{5.46}$$

If we compare (5.45) and (5.42), and choose, in (5.42), $[\omega] = h^3$ as the size of the cells of our μ-space partition, we find that they differ by

$$S_{ST} - S_B = -k(N\log N - N) \approx -k\log N! \tag{5.47}$$

where we have once again used Stirling's approximation. The properly extensive quantity S_{ST} is obtained by subtracting $k\log N!$ from S_B as obtained by a straightforward classical calculation, or, equivalently,

$$S_{ST} = k\log\left(\frac{\Omega}{N!}\right), \tag{5.48}$$

corresponding to a volume of a reduced Γ-space in which states that differ only by permutations of the molecules are identified. Textbooks, (*e.g.* Huang (1986, §6.6)) often say that the justification of this must be quantum-mechanical, though others, (*e.g.* Allis and Herlin (1952); Dugdale (1996)) do a classical calculation with indistiguishability built in from the start. For discussion, see Saunders (2006).

6 Probability

6.1 Enter Probability

A system that, if out of equilibrium, approaches equilibrium, and, if in equilibrium, stays there, will be said to be exhibiting *thermodynamic behavior*. A system moving *away* from equilibrium will be said be exhibiting *anti-thermodynamic behaviour*.

The Reversibility and Recurrence objections to the H-theorem show , respectively,

- We won't see thermodynamic behaviour for all initial states of a gas. In fact, there is a one-one mapping (velocity reversal) between non-equilibrium states that move toward equilibrium, and ones that move away.

- A gas in a box, if isolated for a long period of time, will eventually exhibit anti-thermodynamic behaviour.

These facts, together with our earlier conclusion (§4.2) that it is the revised, probabilistic version of the Second Law that is worth recovering, strongly suggest a revised, probabilistic version of the H-theorem, along the lines that, for macroscopic systems (*i.e* large numbers of degrees of freedom), H will *probably* decrease to a minimum, and stay there for any time period that we are likely to be observing the system (recall that Poincaré recurrence times for large fluctuations are mind-numbingly long). That the H-theorem be given a statistical interpretation was suggested by the Ehrenfests (Ehrenfest and Ehrenfest, 1990). And the conclusion, in the previous section, that states corresponding to macroscopic equilibrium occupy the preponderance (with respect to a measure on phase-space that is uniform over the energy-surface) of phase space, is certainly suggestive of a probabilistic interpretation. But some work must be done to go from "greater phase-space volume" to "more probable." Among other things, we will have to ask what we will *mean* by "more probable."

6.2 Probability: chance and credence

6.2.1 Axioms of probability

A probability function $Pr(\cdot)$ maps propositions to real numbers. The domain of our probability function is assumed to be closed under Boolean combinations. These functions are held to satisfy the following axioms, adapted from Kolmogorov.

1. For all p, $Pr(p) \geq 0$.
2. If p is logically true, $Pr(p) = 1$.
3. If p and q are logically incompatible, then $Pr(p \vee q) = Pr(p) + Pr(q)$.

There is another axiom, adopted by some and omitted by others. We first need an extra assumption about our set of propositions, which is that, for any sequence $\{p_i\}$ of mutually incompatible propositions, there is a proposition, which we will denote $\bigvee_i p_i$, that is true iff one of $\{p_i\}$ is true. The fourth axiom is called *countable additivity*.

4. For any sequence $\{p_i\}$ of mutually incompatible propositions,

$$Pr(\bigvee_i p_i) = \sum_i Pr(p_i).$$

6.2.2 Conditional probability

Given p, q, with $Pr(q) \neq 0$, we define the *conditional probability*

$$Pr(p|q) =_{df.} \frac{Pr(p\&q)}{Pr(q)}. \tag{6.1}$$

Note that this has the consequence that

$$Pr(p\&q) = Pr(p|q)Pr(q) = Pr(q|p)Pr(p), \tag{6.2}$$

which entails that

$$Pr(p|q) = \frac{Pr(q|p)\ Pr(p)}{Pr(q)}. \tag{6.3}$$

This is known as *Bayes' theorem*.

6.2.3 Distinct senses of "probability"

As Hacking (1975) and others have pointed out, the word "probability" is used to cover (at least) two distinct concepts. One concept, the *epistemic* concept, has to do with degrees of belief of a rational agent. The other concept, which may appropriately be called the *aleatory* concept, is the concept appropriate to games of chance; this is the sense in which one speaks, for example, of the probability (whether known or not) of rolling at least one pair of sixes, in 24 throws of a pair of fair dice. I will use the word *credence* for the former, epistemic concept, and *chance* for the aleatory concept. Chances are to be thought of as objective features of a *chance set-up*, such as a coin toss. They are, therefore, the sorts of things one can have degrees of belief about; we have credences about the values of chances.

It can be argued (see, *e.g.*, Greaves and Wallace (2006)) that a rational agent, upon learning a new item of evidence e, will update her credences by conditonalizing on e. So, for any hypothesis h,

$$Cr(h) \to Cr(h|e) = \frac{Cr(e|h)\, Cr(h)}{Cr(e)}. \tag{6.4}$$

6.2.4 Evidence about chances

Suppose that you are about to toss a coin. You initially have degrees of belief about the value of the chance of heads on each toss. Let's pretend that there is a discrete set $\{\lambda_i\}$ of candidate values for the chance (this is for simplicity of exposition only; it's easy enough to lift this restriction). Let h_i be the hypothesis that the chance of heads on each toss is λ_i.

Toss the coin 10 times, and get some result, *e.g.*

$$THHHTTHHHT$$

Let e be this result. The chance of this happening, if h_i is true, is

$$ch(e|h_i) = \lambda_i^6(1-\lambda_i)^4. \tag{6.5}$$

We want to update our credences in the hypotheses $\{h_i\}$ by conditionalizing on e

$$Cr(h_i) \to Cr(h_i|e) = \frac{Cr(e|h_i)\, Cr(h_i)}{Cr(e)} = \frac{Cr(e|h_i)\, Cr(h_i)}{\sum_j Cr(e|h_j) Cr(h_j)}. \tag{6.6}$$

We introduce a condition on rational credence, called by David Lewis (1980) the *Principal Principle*, that constrains the values of $Cr(e|h_i)$.

PP. If h entails that the chance of e is some number x, and b contains only 'admissible' information,[19] then

$$Cr(e|h\&b) = x = ch(e|h). \tag{6.7}$$

The PP is a requirement that our agent's degree of belief in e meshes appropriately with her degrees of belief about the candidates for the chance of e.

$$Cr(e|b) = \sum_i ch(e|h_i)\, Cr(h_i|b). \tag{6.8}$$

If the agent's credence satisfy the Principal Principle,

$$Cr(h_i|e) = \frac{ch(e|h_i)\, Cr(h_i)}{Cr(e)}. \tag{6.9}$$

or,

$$\frac{Cr(h_i|e)/Cr(h_i)}{Cr(h_j|e)/Cr(h_j)} = \frac{ch(e|h_i)}{ch(e|h_j)}. \tag{6.10}$$

The evidence e favours hypotheses that endow e with high chance. For any sequence s of N coin tosses having n heads and $m = N - n$ tails,

$$ch(s|h_i) = \lambda_i^n (1 - \lambda_i)^m. \tag{6.11}$$

It is easy to check that this function achieves a maximum when λ_i is equal to the observed relative frequency of heads, n/N, and is more sharply peaked, the larger N is. This means that, for large N, the chance is high that the relative frequency n/N will be close to λ. In fact, it can be proven (this is Bernoulli's theorem) that, for an infinite sequence of flips, the chance that the relative frequency converges to λ is equal to 1.

A long run of repeated coin flips will strongly favour hypotheses that posit a chance of heads on each flip that is close to the observed relative frequency. Credence, chance, and frequency are linked in an evidential relationship: relative frequency data from repeated experiments is evidence relevant to our credences about single-case chances.

[19] Lewis: "Admissible propositions are the sort of information whose impact on credence about outcomes comes entirely by way of credence about the chances of those outcomes." In particular, e itself is not admissible.

6.3 Frequentism

Consider:

- If an urn contains N balls, of which m are black, and a ball is drawn in such a way that each ball has an equal chance of being the one drawn, then the chance that the ball drawn is black is equal to m/N, the relative frequency of black balls in the urn.
- In an infinite series of Bernoulli trials (this means: successive trials independent, with chance of each possible outcome same on each trial), the chance is equal to one that the relative frequency of outcome x will converge as $n \to \infty$ (and, moreover, will converge to the chance of x in each individual trial).
- Relative frequency data can provide evidence about values of chances.

There are, therefore, intimate connections between chances and relative frequencies. This (together, I think, with worries about what objective chance could mean in a deterministic world), has suggested to some that the "probability", in the objective sense, be identified with relative frequencies in either actual or hypothetical sequences of events. This became a popular idea in the 19th century. John Venn's book *The Logic of Chance* (1866) is one source of this view. An influential 20th-century proponent of the view was Ludwig von Mises, who defended a view of probabilities as frequencies in hypothetical infinite sequences he called *Kollektiv*s.

Most (but not all) philosophers these days think that a frequency interpretation of probability is untenable. The intuition underlying von Mises, that frequencies would converge if certain trials were continued indefinitely, seems to be based on the Bernoulli theorem and related convergence theorems, whose very statement requires a notion of chance independent of frequency: they say that, *with chance equal to one*, relative frequencies will converge. Similarly, the conclusion that the probability that a ball drawn from an urn be black is equal to the relative frequency of black balls in the urn requires some notion that each ball have an equal chance of being drawn. And there is the problem of the *reference class*: if I want to identify an objective probability that a certain item have property P with the relative frequency of P in some class of similar items, this will vary depending on the choice of the class to which the probability is referred. Furthermore, if all I know is that the finite sequence of coin tosses I'm about to see is an initial segment of an

infinite sequence in which the relative frequency of H converges to 1/2; this tells me *nothing at all* about the finite initial sequence. Why? Because, if I take an infinite sequence in which the frequency of heads converges to 1/2, and tack on *any* arbitrary initial finite sequence, no matter how long, then the new sequence has exactly the same convergence properties as the old.

6.4 Classical Probability and Symmetries

Classical probability theory is mostly concerned with situations, approximated by gambling set-ups, in which there is a partition of possible outcomes into equally probable classes (think dice tosses, coin tosses, roulette wheels). Laplace used this as the basis for probability theory, in his seminal *Philosophical Essay on Probability* (1814). Therein he wrote,

> The curve described by a simple molecule of air or vapor is regulated in a manner just as certain as the planetary orbits; the only difference between them is that which comes from our ignorance.
>
> Probability is relative, in part to this ignorance, in part to our knowledge. We know that of three or a greater number of events a single one ought to occur; but nothing induces us to believe that one of them will occur rather than the others. In this state of indecision it is impossible for us to announce their occurrence with certainty. It is, however, probable that one of these events, chosen at will, will not occur because we see several cases equally possible which exclude its occurrence, while only a single one favors it.
>
> The theory of chance consists in reducing all the events of the same kind to a certain number of cases equally possible, that is to say, to such as we may be equally undecided about in regard to their existence, and in determining the number of cases favorable to the event whose probability is sought. The ratio of this number to that of all the cases possible is the measure of this probability, which is this simple a fraction whose numerator is the number of favorable cases and whose denominator is the number of all the cases possible.

Note that there is a qualification: we must judge the cases to be *equally possible*, which seems to be synonymous with "equally probable"[20] If this is what it means, then we can't take this, on pain of circularity, as a *definition* of probability. What Laplace can be thought of as doing is: showing us how, *given* judgments of equiprobability, to get other probabilities out of these judgments.

There is a temptation to think that a judgment of which events are equally probable is dispensable, and that we can define probability as ratio of favourable cases to total number of cases. In fact, Laplace first does this, when he is laying down principles, then corrects himself:

> *First Principle.*—The first of these principles is the definition itself of probability, which, as has been seen, is the ratio of the number of favorable cases to that of all the cases possible.
>
> *Second Principle.*—But that supposes the various cases equally possible. If they are not so, we will determine first their respective probabilities, whose exact appreciation is one of the most delicate points of the theory of chance.

Yes, indeed! There has been extracted from Laplace a *Principle of Indifference*, which says

> If you have no information about a set of mutually exclusive possibilities, assign them equal probabilities.

This can lead to absurd results if applied incautiously. Consider a coin about to be tossed twice. The natural choice of equiprobable partition is $\{HH, HT, TH, TT\}$, each of which is to be assigned equal probability $1/4$. But someone might reason: there are three possibilities for the number of heads, and each of these is equiprobable.

In this case, there is a natural-seeming choice of equipartition. Things get more complicated, though, when there is a continuum of possibilities. There might not be a natural choice for a variable that is uniformly distributed. Consider, for example, a number x to be chosen from the interval $[0, 1]$. This is equivalent to a choice of x^2, as, on this interval the mapping $x \leftrightarrow x^2$ is one-one. Consider: what is the chance that $x > 1/2$? If this chance is $1/2$, then there is chance $1/2$ that $x^2 > 1/4$; and so the chance that $x^2 > 1/2$ is less than $1/2$.

[20] See Hacking (1971) for discussion.

Say that a probability distribution is *uniform* in x, on the interval $[0, 1]$, if for any subinterval $\Delta \subseteq [0, 1]$, $Pr(x \in \Delta)$ is equal to the length of Δ. Then a probability distribution that is uniform in x is not uniform in x^2, and *vice versa*. This matters for statistical mechanics, because, if we want to talk about probabilities being uniform over a system's state space, we have to know: uniform in what variables? Any argument that an appropriate probability distribution for systems in equilibrium is one that is uniform over energy surfaces, *i.e.* uniform in the phase space variables $\{(\mathbf{x}, \mathbf{p})\}$, will have to invoke (or so it seems) some sort of physical considerations, involving the dynamics of the system.

If I know something about symmetries that a set of objective chances satisfy, then this knowledge can be useful. For example, if I know that chances are invariant under a swap $H \leftrightarrow T$, then I know that $ch(H) = ch(T) = 1/2$. More on symmetry reasoning about chances, later. Also, see, if you're interested, Ch. 12 of (van Fraassen, 1989).

6.4.1 Knowledge from ignorance?

Laplace's use of 'probability' seems to be ambiguous between chance and credence. And this poses a danger. If we don't distinguish between chance and credence, calling them both 'probability,' there is a risk of sliding into a fallacy (not, I must say, one that Laplace can be saddled with). Suppose I know nothing about a die, except that it looks symmetrical to me. I know of no reason why one face will be favoured over another. I therefore, have equal credence in each face being the one to land face up. There is a danger of sliding from this, into a claim that I am *certain* that the chances are equal. And this is a very strong claim, not one derivable from ignorance! A judgment of equal chance is not an absence of judgment.

Something like this seems to be going on in, *e.g*, Jackson (1968). See pp. 8-9, 83. (Quoted in Myrvold (2012b)).

6.5 Measure spaces and measures

A *measurable space* is a set X, together with a set $\mathcal{S}$ of subsets of X (the *measurable subsets* of X), such that:

1. $X \in \mathcal{S}$
2. For every $S \in \mathcal{S}$, its complement $X \setminus S$ is also in $\mathcal{S}$.
3. If $\{A_i \ : i \in \mathbb{N}\}$ is a sequence of sets in $\mathcal{S}$, their union, $\bigcup_i A_i$, is also in $\mathcal{S}$.

These conditions define a *σ-algebra*. A σ-algebra that is often of particular interest, when X is endowed with a natural topology, is comprised by the *Borel subsets* of X: the smallest σ-algebra containing all the open sets.

Not all subsets of $\mathbb{R}^n$ are Borel sets, hence not all will be assigned a measure on the measures with which we will be concerned, but this will make little difference.

A *measure* on a measurable space is a function $\mu : \mathcal{S} \to \mathbb{R}^+$ such that:

1. $\mu(\emptyset) = 0$.
2. If $\{A_i\}$ is a sequence of mutually disjoint sets in $\mathcal{S}$,

$$\mu(\bigcup_i A_i) = \sum_i \mu(A_i). \tag{6.12}$$

A measure is *unital* iff $\mu(X) = 1$.

Suppose I have a physical system whose state is represented by a point x in a state space X. Then, for every subset $S \subseteq X$, there will be a corresponding proposition $x \in S$.

If $\langle X, \mathcal{S} \rangle$ is a measurable space, then every unital measure μ defines a probability function, defined on propositions of the form $x \in S$, where $S \in \mathcal{S}$.

$$Pr(x \in S) = \mu(S). \tag{6.13}$$

6.6 Probability flow; Liouville's theorem

Suppose that, at some time t_0, the probability (be it chance or credence) that the state of our system is in $S \subseteq X$ is given by the measure $\rho_0(S)$. Let the system evolve to time t_1. Each point in S will evolve to a new point. Let $U(S)$ be the set of points that S evolves into. Then, the probability that, at time t_1, the system is in the set $U(S)$, is the same as the probability that it was in set S at time t_0. We can define a new probability measure on the phase space by

$$\rho_1(U(S)) = \rho_0(S), \tag{6.14}$$

or,

$$\rho_1(S) = \rho_0(U^{-1}(S)). \tag{6.15}$$

Since we can do this for any time, we can define a family of probability measures, $\rho(t)$, and we can picture the action of the dynamical evolution of the system as a "probability flow."

One defines a *dynamical system* as a quadruple $\langle X, \mathcal{S}, \mu, \phi_t \rangle$, where X is a state space, $\mathcal{S}$ a σ-algebra (the measurable subsets of X), μ a unital measure on $\mathcal{S}$, and $\phi_t : X \to X$ a semi-group of maps from X into itself, representing the dynamical evolution of the system.

You may have encountered, in continuum mechanics, a continuity equation for the flow of a fluid. A fluid with density $\rho(\mathbf{x}, t)$ and velocity-field $\mathbf{v}(\mathbf{x}, t)$ satisfies,

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho\, \mathbf{v}) = 0. \tag{6.16}$$

(This is because the rate of change of the amount of the fluid in any bounded volume must be equal to the rate of flow across the surface of the volume. Apply the Divergence Theorem.)

Our probability density ρ is defined on phase space Γ, which is parameterized by the full set of coordinates and conjugate momenta for all degrees of freedom of the system: $\{(q_i, p_i) \ : i = 1, ..., \dim(\Gamma)/2\}$. The continuity equation for our probability flow is therefore,

$$\frac{\partial \rho}{\partial t} + \sum_i \frac{\partial}{\partial q_i} (\rho\, \dot{q}_i) + \sum_i \frac{\partial}{\partial p_i} (\rho\, \dot{p}_i) = 0. \tag{6.17}$$

Suppose, now, that the system obeys Hamilton's equations of motion,

$$\dot{q}_i = \frac{\partial H}{\partial p_i} \qquad \dot{p}_i = -\frac{\partial H}{\partial q_i}. \tag{6.18}$$

This gives us *Liouville's theorem*

$$\frac{\partial \rho}{\partial t} + \sum_i \left(\frac{\partial \rho}{\partial q_i} \frac{\partial H}{\partial p_i} - \frac{\partial \rho}{\partial p_i} \frac{\partial H}{\partial q_i} \right) = 0. \tag{6.19}$$

Apply this to the measure that is uniform over phase space: $\rho(\mathbf{q}, \mathbf{p}) = const.$. Then $\partial\rho/\partial q_i = \partial\rho/\partial p_i = 0$, and so

$$\frac{\partial \rho}{\partial t} = 0. \tag{6.20}$$

The uniform measure on phase space is invariant under dynamical evolution. Similarly, any measure that is uniform on energy surfaces is invariant. Suppose that ρ can be represented by a density function that is a function of the Hamiltonian: $\rho(\mathbf{q}, \mathbf{p}) = f(H(\mathbf{q}, \mathbf{p}))$. Then

$$\frac{\partial \rho}{\partial q_i} \frac{\partial H}{\partial p_i} - \frac{\partial \rho}{\partial p_i} \frac{\partial H}{\partial q_i} = f'(H) \left(\frac{\partial H}{\partial q_i} \frac{\partial H}{\partial p_i} - \frac{\partial H}{\partial p_i} \frac{\partial H}{\partial q_i} \right) = 0.$$

7 Probabilities in Statistical Mechanics

7.1 The brute posit.

One approach to probabilities in statistical mechanics (see *e.g.* Tolman (1938)) is to simply include, as a postulate of the theory, that the probability of a system being in a certain microstate, given macroscopic constraints, is uniform on energy surfaces consistent with the microstate. This gives rise to the most commonly used probability distributions:

- The *microcanonical distribution.* Appropriate for an isolated system, which has a fixed energy. Take a region of phase space containing all states with energy in a narrow band $[E, E+\delta E]$, and take a probability distribution that is uniform, in canonical phase space variables, in this small region. One can also define a probability distribution on the surface of energy E by taking a limit.

- The *canonical distribution.* Appropriate for systems in contact with a heat bath of fixed temperature (the system can exchange energy with the heat bath, and hence its energy can vary.) Given by

$$\rho(\mathbf{q}, \mathbf{p}) \propto e^{-\beta \mathcal{E}(\mathbf{q},\mathbf{p})}, \tag{7.21}$$

  where $\mathcal{E}(\mathbf{q}, \mathbf{p})$ is the energy of the system in state $(\mathbf{q}, \mathbf{p})$, and, as before, $\beta = 1/kT$. Because physicists, from the time of Gibbs, have tended to visualize probability distributions as large collections of systems subjected to the same preparation procedure, the terminology *microcanonical ensemble* and *canonical ensemble* is widespread.

As these are stationary distributions, the expectation value ('ensemble average') of any state function will be constant in time for such probability distributions. Consequently, Gibbs (and following him, many of the standard textbooks) identified equilibrium, not with a particular state of a system, but with such a probability distribution over states of the system.

The textbooks tend to adopt a frequency approach to probability, which was criticized in §5.3. One might, alternatively, interpret these probabilities either as ideal credences, or as chances. On the former interpretation, they represent the state of knowledge of someone who knows only the macrostate. On the latter, we might think of typical preparations of systems in a given

thermodynamic state as chance set-ups, with which are associated chances of yielding the possible microstates.

Suppose that we adopt one of these interpretations, and apply it to our good old example of an isolated gas initially confined to one side of a container. If we form our expectations about its initial state, and hence about what will happen when the partition is removed, on the basis of the microcanonical distribution, then we find (since our credences are equally distributed over all velocity directions for each molecule), that we will expect the gas to fill the container, at pretty much the rate that it does. Similar remarks for a chance interpretation: on the microcanonical distribution, the chance is high that the gas will fill the box, low that it will do anything else.

So far so good! Suppose, now, that we let the gas relax to its new equilibrium state, and apply the microcanonical distribution adapted to its new macroscopic constraints. The expectations we form on the basis of this regarding subsequent behaviour, including fluctuations, seem to fit our experience very well. If we take seriously the idea that this probability distribution represents our state of knowledge, however, there is a problem. This is most keenly seen if we use the probability distribution as a basis for retrodictions. The microcanonical distribution knows no direction of time; we expect, on its basis, the same sort of fluctuations to the past that we expect to the future. But we know something about the state that has gotten lost: namely, that it recently was in only one half of the container (and hence, if the dynamics are deterministic and TRI, that a time-reversal of the trajectories of all the molecules would lead them back into their original volume). David Albert has put this in a characteristically emphatic manner.

> this latest version of the statistical postulate, if applied in the present, is *flatly inconsistent* with what we take to be true, with what we *remember*, with what is *recorded* ... of the *past*. (Albert, 2000, p. 81)

Albert proposes instead, that we adopt, as our postulate about probabilities, a probability distribution that is uniform over the part of phase space consistent with the known macroscopic state of a system and what he calls the *Past Hypothesis*, "which is that the world first came into being in whatever particular low-entropy highly condensed big-bang sort of macrocondition it is that the normal inferential procedures of cosmology will eventually present to us" (Albert, 2000, p. 96).

Note: the claim is merely that this prescription yields reasonable inference, about both the past and the future. No explanation is being offered as to *why* systems exhibit thermodynamic, rather than anti-thermodynamic, behaviour. Even this relatively modest claim has been disputed; Winsberg (2004) argues that it does not succeed in yielding the right inferences about the recent past of small, temporarily isolated systems, and Earman (2006) has questioned whether it makes sense to talk about the entropy of the early Universe.

7.2 Appeals to typicality

There is a picture that underlies much of the contemporary discussion of time asymmetry in statistical mechanics. It is articulated very clearly in Price (1996), and is found also in, *e.g.* Goldstein (2001), Penrose (1990). Ultimately, it has its origins in Boltzmann.

The picture is this: as we have seen, high-entropy states occupy the majority of the phase space accessible to a system. We should therefore regard such states as 'typical,' and expect to find systems in high-entropy, that is, equilibrium states. If a system is found in a low-entropy, non-equilibrium state, then, since this is an unusual state, we should not be surprised if it wanders into a more typical region of phase space. I quote Huw Price:

> The history of science shows that science often changes our conception of what calls for explanation and what doesn't. Familiar phenomena come to be seen in a new light, and often as either more or less in need of explanation as a result. One crucial notion is that of normalcy, or naturalness. Roughly, things are more in need of explanation the more they depart from their natural condition. (The classic example is the change that Galileo and Newton brought about in our conception of natural motion.)
>
> ... Thus it seems to me that the problem of explaining why entropy increases has been vastly overrated. The statistical considerations suggest that a future in which entropy reaches its maximum is not in need of explanation; and yet that future, taken together with the low-entropy past, accounts for the general gradient. (Price, 1996, pp. 39–40)

On this view, an increase of entropy should be regarded as natural, not standing in need of explanation. The reasoning that leads to this, however,

is entirely time-symmetric. What is puzzling, on this view, is why entropy *doesn't* increase towards the past. What calls for explanation is the low-entropy early universe.

... But something has gone wrong here. Recall that Boltzmann's original H-theorem sought to explain the relaxation to equilibrium by consideration of molecular collisions. We now seemingly have an explanation of relaxation to equilibrium that dispenses with dynamical considerations: the move from a non-equilibrium macrostate to an equilibrium macrostate is just a move from an atypical region of phase space to a more typical region, a move from an 'unnatural' state to a more 'natural' one, a move that does not stand in need of explanation. That something important has been left out can been seen by considering, once again, a gas confined to one side of a box. A higher entropy state would be one in which the gas is evenly distributed on both sides of the box. But the mere fact that it can thereby increase its entropy isn't going to permit the gas to seep through the partition; it remains confined to one side until the partition is removed.

Now remove the partition. The conclusion that the gas does not remain in a low-entropy state, and approaches the maximum-entropy state compatible with the new set of macroscopic constraints, must be based on *some* assumption about its dynamics. Suppose, for example, that there is some constant of the motion that we've forgotten to mention, such that not all regions of the phase space are accessible from its initial state, and among the inaccessible regions are the ones of maximum entropy. The Pricean argument carries an implicit assumption that this isn't the case, and, if this assumption is correct, then it must be based on some feature of the system's dynamics.

7.3 The Ergodic Hypothesis

Boltzmann conjectured that

> The great irregularity of the thermal motion and the multitude of forces that act on a body make it probable that its atoms, due to the motion that we call heat, traverse all positions and velocities which are compatible with the principle of [conservation of] energy (quoted in Uffink (2007b, 40)).

This has come to be known as the *ergodic hypothesis*. As stated, it cannot be correct, as the trajectory is a one-dimensional continuous curve and so cannot fill a space of more than one dimension. But it *can* be true that

almost all trajectory eventually enter every open neighbourhood of every point on the energy surface. Boltzmann argued, on the basis of the ergodic hypothesis, that the long-run fraction of time that a system spends in a given subset of the energy surface is given by the measure that Gibbs was to call microcanonical.

Given a Hamiltonian dynamical system, and an initial point x_0, we can define, for any measurable set A such that the requisite limit exists, the quantity

$$\langle A, x_0 \rangle_{time} = \lim_{T \to \infty} \frac{1}{T} \int_0^T \chi_A(T_t(x_0))\, dt, \tag{7.22}$$

where χ_A is the indicator function for A,

$$\chi_A(x) = \begin{cases} 1, & x \in A \\ 0, & x \notin A. \end{cases} \tag{7.23}$$

$\langle A, x_0 \rangle_{time}$, provided it exists, is the fraction of time, in the long run, that a trajectory starting at the point x_0 spends in the set A.

A dynamical system is said to be *ergodic* iff, for any set A such that $\mu(A) > 0$, the set of initial points that never enter A has zero measure. It is easily shown that this condition is equivalent to *metric transitivity*: a dynamical system is *metrically transitive* iff, for any partition of Γ into disjoint subsets A_1, A_2 such that, for all t, $T_t(A_1) \subseteq A_1$ and $T_t(A_2) \subseteq A_2$, either $\mu(A_1) = 0$ or $\mu(A_2) = 0$.

Von Neumann and Birkhoff proved that, for any measure-preserving dynamical system,

1. For any $A \in \mathcal{S}$, the limit

$$\langle A, x_0 \rangle_{time} = \lim_{T \to \infty} \frac{1}{T} \int_0^T \chi_A(T_t(x_0))\, dt \tag{7.24}$$

   exists for almost all points x_0. (That is, if X is the set of points for which this limit doesn't exist, $\mu(X) = 0$.)

2. If the dynamical system is ergodic, then

$$\langle A, x_0 \rangle_{time} = \mu(A). \tag{7.25}$$

   for all $A \in \mathcal{S}$ and almost all $x_0 \in \Gamma$.

How to go from this measure to probability? If a system, isolated for a long time, is observed at some random time, then the chance of finding it in some region C of its phase space is equal to $\mu(C)$. This gives us a sense of the long-run behaviour of an isolated system: most of the time it is at or near thermodynamic equilibrium, with occasional fluctuations away from equilibrium, and, in the long run, spends as much time moving towards equilibrium as moving away. Large fluctuations will be much rarer than small ones. This means that, if we look at an isolated system at a randomly selected time t and find $H > H_{min}$ ($S < S_{max}$), and ask whether the value of H will be higher or lower some at sometime $t + \delta t$, a short time afterward, then it is more probable that H will be lower (S will be higher) at $t + \delta t$.

Is this a justification for taking the microcanonical measure to be the measure that yields the correct probabilities for an isolated system? Two reservations arise.

The first is the question whether actual systems of interest have ergodic dynamics. Proving this turns out to be very difficult even for some very simple systems. Moreover, there are systems, namely, those to which the KAM theorem applies, that are provably not ergodic (see Berkovitz et al. (2006) for discussion of the applicability of ergodic theory).

The second is the use of the long-term time average. The picture invoked above, of a system isolated for a very long time and observed at a random time, does not fit neatly with laboratory procedures. One argument that has been given for considering the long-term time average is as follows.[21] Measurements of thermodynamic variables such as, say, temperature, are not instantaneous, but have a duration which, though short on human time scales, are long on the time scales of molecular evolution. What we measure, then, is in effect a time-average over a time period that count as a very long time period on the relevant scale.

This rationale is problematic. The time scales of measurement, though long, are not long enough that the average over them necessarily approximates the limit in (7.22); as Sklar (1993, 176) points out, if they were, then the only measured values we would have for thermodynamic quantities would be equilibrium values. This, as Sklar puts it, is "patently false"; we are, in fact, able to track the approach to equilibrium by measuring changes in thermodynamic variables.

[21] Adapted from Khinchin (1949, 44-45).

Another issue, raised by Sklar among others, is the "measure-zero" problem. The ergodic theorem gives us conclusions valid for all but a set of initial points having measure zero. We would like to say that we can neglect the exceptional points. Can a non-circular justification be given? For an insightful discussion, see Malament and Zabell (1980).

As mentioned above, if we are to ask for a probability distribution appropriate to thermodynamic equilibrium, the distribution should be a stationary distribution. The microcanonical distribution is a stationary distribution on Γ_E. If the system is ergodic, then it is the *only* stationary distribution among those that assign probability zero to the same sets that it does. For a justification of the use of the microcanonical distribution along these lines, see Malament and Zabell (1980).

7.4 Boltzmann-Schuetz cosmology

We concluded, above, that, if we observe an ergodic system at a random time, and find it in a state of nonmaximal entropy, then the entropy S is more likely to be greater, a short time δt later, than it is to be smaller. This looks like a temporally asymmetric conclusion, but it's not. Since small fluctuations are so much more common then larger ones, if we observe the system and see a state that is not the equilibrium state, we should conclude that we are probably near a local minimum of entropy. We should, therefore, conclude, not only that the entropy is likely to be higher in the near future, but also that it is likely that it was higher in the near past.

In a letter to *Nature* in 1895, Boltzmann applied considerations of this sort to the Universe as a whole, crediting the idea to his assistant Schuetz.[22]

> If we assume the universe great enough, we can make the probability of one relatively small part being in any given state (however far from the state of thermal equilibrium), as great as we please. We can also make the probability great that, though the whole universe is in thermal equilibrium, our world is in its present state. It may be said that the world is so far from thermal equilibrium that we cannot imagine the improbability of such a

[22]My conjecture, in a previous version of these notes, that the name should be "Schütz," and that "Schuetz" was an accommodation to English typesetting, turns out, on closer inspection, to be untenable— in the same letter, umlauts in titles of papers that Boltzmann refers to are handled without a problem.

> state. But can we imagine, on the other side, how small a part of the whole universe this world is? Assuming the universe great enough, the probability that such a small part of it as our world should be in its present state, is no longer small. (Boltzmann, 1895)

This conjures up a vision of a time-symmetric cosmology, with most of the universe in a state of thermodynamic equilibrium, occasional small fluctuations, rarer large fluctuations, one of which is our home. There would also, no doubt, be, in other parts of the universe, separated by a vast sea of equilibrium matter, beings whose arrow of time is oppositely directed from ours. (See also Boltzmann (1995, §90).)

One attractive feature of this view is that there is no genuine temporal asymmetry to be explained. There are various local asymmetries, but no temporal direction distinguished overall.

> Very well, you may smile at this; but you must admit that the model of the world developed here is at least a possible one, free of inner contradiction, and also a useful one, since it provides us with many new viewpoints. (Boltzmann, 1995, p.448)

The Boltzmann-Schuetz cosmology is a weird idea, but weirdness *alone* ought not be held as an objection.

There is a consequence, however, that Boltzmann seems not to have noticed. On such a scenario, the vast majority of occurrences of a given non-maximal level of entropy would be near a local entropy minimum, and so one should regard it as overwhelmingly probable that, even given our current experience, entropy increases towards the past as well as the future, and everything that seems to be a record of a lower entropy past is itself the product of a random fluctuation. Moreover, you should take yourself to be whatever the minimal physical system is that is capable of supporting experiences like yours; apparent experiences of being surrounded by an abundance of low-entropy matter are illusory. That is, you should take yourself to be what has been called a "Boltzmann brain."[23]

[23]The term is due to Andreas Albrecht. It first appears in print in Albrecht and Sorbo (2004). The consequence of the Boltzmann-Schuetz cosmology, that we should take the fluctuation we are in to be no larger than necessary, seems to have been first pointed out by Arthur Eddington.

Incredible? Yes. But, if the universe is big enough, and lasts long enough, all sorts of chance fluctuations are possible.

But there's a problem here. If anything like this story is correct (and, as Boltzmann says, it's not self-contradictory), then what I take to be records of experiments (and memories of reading about them), are "forgeries"—those experiments didn't happen. But we were led to this picture because we wanted to take statistical mechanics seriously, and at least part of our reason for taking statistical mechanics seriously is its alleged empirical success—which now seems to be bogus. The Boltzmann-Schuetz cosmology is in that peculiar class of theories that are *empirically self-undermining.*[24] Though it is not impossible for the theory to be true, the theory itself tells us to disbelieve the sorts of things that we might be tempted to take as empirical evidence for the theory. Clearly, something has gone seriously wrong.

7.5 Almost-objective probabilities

There is, in the mathematical literature on probability, a family of techniques that is known (somewhat misleadingly) as "the method of arbitrary functions." The idea is that, for certain systems, a wide range of probability distributions will be taken, via the dynamics of the system, into distributions that yield approximately the same probabilities for some statements about the system.[25]

As an example, consider a variant on Poincaré's wheel[26] A roulette-like wheel is divided into a large, even number n of equal sectors, alternately

> (unless we admit something which is not chance in the architecture of the universe) it is practically certain that a universe containing mathematical physicists will at any assigned date be in the state of maximum disorganization which is not inconsistent with the existence of such creatures (Eddington, 1931, 452).

The implication one should take oneself to be a brain, rather than entire mathematical physicist, was drawn by Martin Rees (1997, 221).

[24] I owe this phrase to Barrett (1999).

[25] The method of arbitrary functions was pioneered by von Kries (1886) and Poincaré (1912), and elaborated by a number of mathematicians, notably Hopf (1934, 1936). For a systematic overview of mathematical results, see Engel (1992); for the history, see von Plato (1983).

[26] The wheel is discussed by Poincaré in *Science and Hypothesis* (Ch. XI) and in *Science and Method* (Ch. IV, §VI); his analysis is presented in his lectures on probability (Poincaré 1912). It was treated in considerably more detail by Hopf (1934).

colored red and black. The wheel is spun. A probability function over initial conditions, together with the law of friction that leads the wheel to eventually come to rest, yields a probability distribution over the final resting place of the wheel. If the wheel is spun hard enough so that it undergoes several revolutions before coming to rest, small uncertainties about the initial speed spread to become larger uncertainties about the final resting place of the wheel. Provided the initial distribution is not "too wiggly", and provided that the sectors are small enough, the probabilities assigned to red and black will approximate those assigned by a uniform distribution over the circle.[27]

Though the example involves irreversible dynamics, the result does not depend on this, and one can also consider cases in which the wheel spins freely, without friction, and we ask for its position a specified time interval T after it is spun. Though the dynamics are reversible, and hence an exact specification of the state of the wheel at time T uniquely specifies its initial state (and hence a probability distribution at time T uniquely determines a probability distribution over initial conditions), at the level of coarse-grained observables there is a "forgetting" of initial conditions.

It is plausible, at least, that the dynamics of the sorts of systems to which we successfully apply statistical mechanics exhibit this sort of forgetting of initial conditions. Consider, for example, an isolated system that is initially out of equilibrium (it might, for example, be a cup of hot water with an ice cube in it). It is left alone to relax to equilibrium. Once it has done so, then, it seems, all trace of its former state has been lost, or rather, buried so deeply in the details of the system's microstate that no feasible measurement can be informative about it. For systems of this sort, a wide class of probability distributions over initial conditions evolve, via Liouville's equation, into distributions that, *as far as feasible measurements are concerned*, yield probabilities that are indistinguishable from those yielded by the equilibrium distribution.

We need not restrict ourselves to states of thermodynamic equilibrium. If we open a thermos bottle and find in it half-melted ice cubes in lukewarm water, it is plausible that no feasible measurement on the system will determine whether the system was prepared a few minutes ago with only a little less ice, or an hour ago with boiling water and a lot of ice. If this is right,

[27]For details of the mathematical results, see Engel (1992), and also the discussion in Myrvold (2012c). Here the appropriate measure of wiggliness is the total variation of the density function.

then again, a wide variety of probability distributions over initial conditions will evolve into ones that yield virtually the same probabilities for results of feasible measurements.

Ideas of this sort have recently drawn the attention of philosophers; Strevens (2003, 2011), Rosenthal (2010, 2012), Abrams (2012), and Myrvold (2012a,c) for an array of recent approaches in which the method of arbitrary functions plays a role all invoke the method of arbitrary functions, in different ways.

The method does not generate probabilities out of nothing; rather, the key idea is that a probability distribution over initial conditions is transformed, via the dynamical evolution of the system, into a probability distribution over conditions at a later time. Hence any use of the method must address the question: what is the status of the input distributions? Poincaré describes them as "conventions," which, it must be admitted, is less than helpful. Strevens (2003) is noncommittal on the interpretation of the input probabilities, whereas Strevens (2011) and Abrams (2012) opt for actual frequencies.

Savage (1973) suggested that the input probabilities be given a subjectivist interpretation. For the right sorts of dynamics, large differences in subjective probabilities will lead to probability distributions that agree closely on outcomes of feasible measurements; hence the output probabilities might be called "almost objective" probabilities. This suggestion is developed in Myrvold (2012a,c). The conception combines epistemic and physical considerations. The ingredients that go into the characterization of such probabilities are:

- a class $\mathcal{C}$ of credence-functions about states of affairs at time t_0 that is the class of credences that a reasonable agent could have, in light of information that is accessible to the agent,
- a dynamical map T_t that maps states at time t_0 to states at time $t_1 = t_0 + t$, inducing a map of probability distributions over states at time t_0 to distributions over states at t_1,
- a set $\mathcal{A}$ of propositions about states of affairs at time t_1, to which probabilities are to be assigned,
- a tolerance threshold ϵ for differences in probabilities below which we regard two probabilities as essentially the same.

Given these ingredients, we will say that a proposition $A \in \mathcal{A}$ has an *almost-objective probability*, or *epistemic chance*, if all probability functions in $\mathcal{C}$ yield, when evolved to t_1, essentially the same probability for A. That is, A has epistemic chance λ if, for all $P_0 \in \mathcal{C}$, $|P_t(A) - \lambda| < \epsilon$.

This concept includes an epistemic aspect, as an essential ingredient is the class $\mathcal{C}$ of credence-functions that represent reasonable degrees of belief for agents with our limitations.[28] It would be nice if, for some nontrivial event A, the dynamical map T_t yielded the same probabilities for absolutely all input measures, eliminating the need for a restriction to a class $\mathcal{C}$ of input measures, but this cannot be. However, the physics plays a key role; the value of an epistemic chance, if it exists, is largely a matter of the dynamics.

Those who hold that epistemic considerations ought not to be brought into physics at all will not be happy with construing statistical mechanical probabilities in this way. However, on the Maxwellian view of thermodynamics and statistical mechanics, on which the fundamental concepts of thermodynamics have to do with our ability to keep track of and manipulate molecular motions, this sort of blending of epistemic and physical consideration is just what one would expect to find in statistical mechanics. Epistemic limitations have to do with the abilities of agents; however, what agents with given limitations are able to do with physical systems has to do with the physics of those systems.

7.6 Probabilities from quantum mechanics?

So far, we have been considering classical statistical mechanics. However, our world is not classical; it is quantum. Most writers on the foundations of statistical mechanics have assumed, implicitly or explicitly, that the conceptual problems of classical statistical mechanics are to be solved in classical terms; classical statistical mechanics should be able to stand on its own two feet, as an autonomous science, albeit one that is gets certain facts about the world, such as the specific heats of non-monatomic gases, wrong.

One argument for this might be that we successfully apply statistical mechanics to systems for which quantum effects are negligible. This is questionable. Though, because of the reversibility argument, we know that there are trajectories through phase space that exhibit anti-thermodynamic behaviour, these are unstable under random perturbations. Albrecht and Phillips (2012)

[28] Objective Bayesians would hold that this class is a singleton.

estimate the relevance of quantum uncertainty stock examples such as coin flips and billiard-ball gases, and conclude that "all successful applications of probability to describe nature can be traced to quantum origins."

As emphasized by Albert (2000, Ch. 7), if we consider isolated quantum systems, and assume the usual Schrödinger evolution to be valid at all times, then this leaves us in pretty much the same conceptual situation as in classical mechanics. The dynamics governing the wave-function are reversible; for any state that exhibits the expected thermodynamic behaviour there is a state that exhibits anti-thermodynamic behaviour. Moreover, the von Neumann entropy—the quantum analog of the Gibbs entropy—is conserved under dynamical evolution. Considering nonisolated systems only pushes the problems further out; the state of the system of interest plus a sufficiently large environment can be treated as an isolated system; there will be states of this larger system that lead to antithermodynamic behaviour of the subsystem of interest.

If, however, collapse of the wave-function is a genuinely chancy, dynamical process,then things are different.[29] For any initial state, there will be objective probabilities for any subsequent evolution of the system. Albert has argued that these probabilities suffice to do the job required of them in statistical mechanics.

This is indeed plausible, though we lack a rigorous proof. If this proposal is correct, we should expect that, on time scales expected of relaxation to equilibrium, the probability distribution yielded by the collapse dynamics approaches a distribution that is appropriately like the standard equilibrium distribution, where "'appropriately like" means that it yields approximately the same expectation values for measurable quantities. It is not to be expected that the equilibrium distribution be an exact limiting distribution for long time intervals; in fact, distributions that are stationary under the usual dynamics (quantum or classical) will *not* be strictly stationary under the stochastic evolution of dynamical collapse theories such as the Ghirardi-Rimini-Weber (GRW) or Continuous Spontaneous Localization (CSL) theory, as energy is not conserved in these theories. However, energy increase will be so small as to be under ordinary circumstances unobservable—Bassi and Ghirardi (2003, 310) estimate, for a macroscopic monatomic ideal gas, a temperature increase on the order of 10^{-15} Celsius degrees per year—and so

[29]See Ghirardi (2011); Bassi and Ghirardi (2003) for overviews of dynamical collapse theories.

we might expect relaxation to something closely approximating a standard equilibrium distribution, on the time scales we would expect this to happen, followed by exceedingly slow warming.

7.7 Return of the *Stoßzahlansatz*?

If we don't have a royal road to the conclusion that systems out of equilibrium tend to move toward equilibrium, then we have to do the hard slog of showing this for something like a realistic model. This requires some model of the dynamics, and some assumptions about initial conditions. Boltzmann's H-theorem can be thought of as a first step along this path. It relied, however, on the *Stoßzahlansatz*, which, even if it is a reasonable condition to impose at an initial instant, cannot continue to hold, even approximately, indefinitely far into the future. Some work has been done to model the approach to equilibrium in a more rigorous fashion.

Of note in this vein is the work of Lanford, who obtained an approximation to Boltzmann's transport equation, valid for dilute gases, assuming initial conditions satisfying the equivalent of the *Stoßzahlansatz* . See Uffink (2007a, §6.4) for a lucid exposition.

Also of interest is the work of Bogolyubov, Born, Green, Kirkwood, and Yvon (BBGKY), on the approach of Gibbsian ensembles to equilibrium. Discussion of this is beyond the scope of this course. See Uffink (2007a, §6.5) if you're interested. Of relevance for our purposes: here again, an analog of the *Stoßzahlansatz* is imposed on initial conditions, in the form of an absence of initial correlations.

7.7.1 Is the *Stoßzahlansatz* justified?

The *Stoßzahlansatz*-like assumptions, which amount to absence of correlations, are introduced as initial conditions. They can't continue to hold; as molecules interact with each other, correlations arise between their states. The intuition behind such an assumption seems to be something like a *common cause* principle: Correlations are to be explained on the basis of past interactions.

Here's how Boltzmann put the assumption (which he called the 'hypothesis of molecular disorder,' in the exposition of the H-theorem in his *Lectures on Gas Theory*, written after the discussions about reversibility and recurrence.

> We have a molecular-ordered distribution if—to select only two examples from the infinite manifold of possible cases—each molecule is moving toward its nearest neighbor, or again if each molecule whose velocity lies between certain limits has ten much slower molecules as nearest neighbors. (Boltzmann, 1995, p. 40)

Note that this way of putting it emphasizes the temporally asymmetric nature of the assumption. We don't expect molecules to be disproportionately headed towards their nearest neighbours. To find them disproportionately headed *away* from their nearest neighbours seems less surprising—this will be true of those that have recently undergone collisions.

That we find the *Stoßzahlansatz* reasonable in one temporal direction but not the other has, by some, been taken as an unjustifiable bias on our part. This is one of the chief themes of Price's book (Price, 1996). On the other hand, others seek to justify this on the basis of a temporally asymmetric common-cause principle. On this view, the chief difference between states leading to thermodynamic behaviour, and states leading to anti-thermodynamic behaviour, is that the latter contain correlations not explicable on the basis of a past common cause. In particular, Penrose (2001) defends such a view.

8 Gibbs entropy

8.1 Canonical distribution and entropy

Consider a system that is in thermal equilibrium with a heat bath of temperature T. Suppose that the Hamiltonian of the system contains terms that depend on a number of external parameters $\{\theta_i\}$, which we can manipulate to do work on the system. For example, the walls of a container containing a gas or liquid may be modelled by a potential that is zero except when a molecule is close to the wall, and strongly repulsive at the location of the wall. This term will depend on the position of the wall, and, if the wall is moveable, as is a piston, then we can manipulate this position and compress or expand the gas.

Suppose, further, that we have argued, one way or another, that the appropriate probability distribution for a system in thermal equilibrium with a heat bath of temperature T is the canonical distribution, which is represented by a density function

$$\rho(x) = Z^{-1}\, e^{-\beta H(x)}, \tag{8.1}$$

where $\beta = 1/kT$, and Z is a normalization constant,

$$Z = \int e^{-\beta H(x)} dx. \tag{8.2}$$

The quantity Z, is a function of β and the external parameters θ_i, and is called the *partition function*. It can be used to calculate expectation values of many quantities of interest. Note that

$$\frac{\partial}{\partial\beta} \log Z = Z^{-1}\, \frac{\partial Z}{\partial\beta} = -Z^{-1} \int H(x) e^{-\beta H(x)} dx = -\langle H \rangle, \tag{8.3}$$

or,

$$\langle H \rangle = -\frac{\partial}{\partial\beta} \log Z. \tag{8.4}$$

Similarly,

$$\left\langle \frac{\partial H}{\partial\theta_i} \right\rangle = -\frac{1}{\beta} \frac{\partial}{\partial\theta_i} \log Z. \tag{8.5}$$

Also, since

$$\log \rho(x) = -\log Z - \beta\, H(x), \tag{8.6}$$

we have

$$\begin{aligned}\langle \log \rho \rangle &= -\log Z - \beta \langle H \rangle \\ &= -\log Z + \beta \frac{\partial}{\partial \beta} \log Z \\ &= \beta^2 \frac{\partial}{\partial \beta} \frac{1}{\beta} \log Z. \end{aligned} \tag{8.7}$$

Suppose that we slowly change the parameters θ_i by small amounts $\delta\theta_i$, possibly changing the temperature, let the system settle into a new equilibrium, and put it into contact with a heat bath at the new temperature. There will be a new canonical distribution corresponding to the new parameters and temperature.

Let us investigate the change in $\langle H \rangle$, the expectation value of the energy. This will change not only because of the dependence of H on the parameters θ_i, but also because of the shift of the probability distribution ρ to a canonical distribution appropriate to the new Hamiltonian and temperature.

From

$$\langle \log \rho \rangle = -\log Z - \beta \langle H \rangle. \tag{8.8}$$

we get

$$\langle H \rangle = -\frac{1}{\beta} \langle \log \rho \rangle - \frac{1}{\beta} \log Z. \tag{8.9}$$

Therefore,

$$\delta \langle H \rangle = -\frac{1}{\beta} \delta \langle \log \rho \rangle + \left(\frac{1}{\beta^2} \langle \log \rho \rangle - \frac{\partial}{\partial \beta} \frac{1}{\beta} \log Z \right) \delta\beta - \frac{1}{\beta} \sum_i \left(\frac{\partial}{\partial \theta_i} \log Z \right) \delta\theta_i. \tag{8.10}$$

From (8.7), the middle term of the RHS of (8.10) vanishes. Therefore,

$$\begin{aligned}\delta \langle H \rangle &= -\frac{1}{\beta} \delta \langle \log \rho \rangle - \frac{1}{\beta} \sum_i \left(\frac{\partial}{\partial \theta_i} \log Z \right) \delta\theta_i \\ &= -kT \, \delta \langle \log \rho \rangle + \sum_i \left\langle \frac{\partial H}{\partial \theta_i} \right\rangle \delta\theta_i. \end{aligned} \tag{8.11}$$

Compare this with the First Law of Thermodynamics,

$$dU = đQ + đW. \tag{8.12}$$

For a qsr process we have $đQ = TdS$, and so

$$dU = TdS + đW. \tag{8.13}$$

The last term of (8.11) is the expectation value of the work done on the system by manipulating the parameters θ_i. This suggests that we relate the first term to a heat exchange $đQ = TdS$, which in turn suggests identification of the quantity

$$S_G = -k \langle \log \rho \rangle = -k \int \rho(x) \log \rho(x)\, dx \tag{8.14}$$

with the thermodynamic entropy—at least, for situations in which the canonical distribution is appropriate.

8.2 Properties of the Gibbs entropy

Though we introduced it on the basis of considerations of the canonical distribution, the functional

$$S[\rho] = -k \int_\Gamma \rho(x) \log \rho(x)\, dx \tag{8.15}$$

is well-defined for any probability density function ρ.

Some properties of this quantity:

1. *Invariance.* If ρ evolves according to Liouville's equation, as it will for an isolated system, then $S[\rho]$ is constant in time.

2. *Additivity for Independent Systems.* Let A and B be two disjoint systems, AB the compound system consisting of A and B. If ρ_{AB} is a probability distribution over the state space of the compound systems on which the states of the two component systems are probabilistically independent, that is, if

   $$\rho_{AB}(x_A, x_B) = \rho_A(x_A)\rho_B(x_B), \tag{8.16}$$

   then

   $$S_{AB}[\rho_{AB}] = S_A[\rho_A] + S_B[\rho_B]. \tag{8.17}$$

3. *Subadditivity.* For any probability distribution over the composite system AB,

$$S_{AB}[\rho_{AB}] \leq S_A[\rho_A] + S_B[\rho_B]. \tag{8.18}$$

4. *Concavity.* If ρ_1, ρ_2, are two probability distributions, then

$$S[w\rho_1 + (1-w)\rho_2] \geq wS[\rho_1] + (1-w)S[\rho_2]. \tag{8.19}$$

8.3 Gibbs on Thermodynamic Analogies

The reasoning in the previous section, leading to the suggestion of the identification of S_G—which has come to be called the *Gibbs entropy*—with thermodynamic entropy, is due to Gibbs (1902).[30]

One might worry, though, that (8.11) involves expectation values of change of total energy H and of $\log \rho$, whereas the thermodynamic equation (8.13) involves changes of actual values of state functions U and S. Furthermore, the argument that S_G behaves like an entropy depended on the canonical distribution; should we extend it to situations in which this is not the appropriate probability distributions?

In connection with the first worry, Gibbs notes that it can be shown that, for a canonical distribution for a system with a large number of degrees of freedom, the dispersion in energy will be small compared to the total energy. As Gibbs was able to show, for a system with N degrees of freedom,

$$\frac{\mathrm{Var}(H)}{\langle H \rangle^2} = \frac{\langle (H - \langle H \rangle)^2 \rangle}{\langle H \rangle^2} \propto \frac{1}{N}. \tag{8.20}$$

Thus, for macroscopic systems, the probability that the system's energy is far from its expectation value is small. For Gibbs, the goal of statistical mechanics is "to show by *a priori* reasoning that for such systems as the material bodies which nature presents to us, these [thermodynamic] relations hold with such approximation that they are sensibly true for human faculties of observation" (1902, p. 166). Gibbs continues,

> This indeed is all that is really necessary to establish the science of thermodynamics on an *a priori* basis. Yet we will naturally desire to find the exact expression of those principles of which

[30] See pp. 43–45, and Ch. XIV. Our equation (8.11) is Gibbs' (114) on p. 44, which reappears as (483) on p. 168.

> the laws of thermodynamics are the approximate expression. A very little study of the statistical properties of conservative systems of a finite number of degrees of freedom is sufficient to make it appear, more or less distinctly, that the general laws of thermodynamics are the limit toward which the exact laws of such systems approximate, when their number of degrees of freedom is indefinitely increased.

A bit better way of putting it might be: the general laws of thermodynamics are the behaviour that is approximated with high probability by the behaviour given by the exact laws, when the number of degrees of freedom is indefinitely increased.

What about the worry that the argument that S_G behaves like an entropy depends on using the canonical distribution? This would be a worry if S_G were to be taken quite generally to be the statistical mechanical analog of entropy. But Gibbs does *not* do this. As Gibbs notes, if we are seeking quantities, defined for all N, whose behaviour approximates that of thermodynamical temperature and entropy for large N, then these quantities will not be uniquely determined. "There may be therefore, and there are, other quantities which may be thought to have some claim to be regarded as temperature and entropy with respect to systems of a finite number of degrees of freedom" (1902, p. 169).

Gibbs proceeds to discuss the microcanonical distribution, and identifies quantities other than $1/\beta$ and $\langle \log \rho \rangle$ that correspond to temperature and entropy, respectively, for systems for which this is an appropriate probability distribution, that is, for isolated systems whose total energy is known. This may come as a surprise (it did to me) for those who are used to the customary identification of S_G as *the* Gibbs entropy; the limited scope of the identification is rarely mentioned in the literature. Notable exceptions are Uffink (2007a) and Batterman (2010).

8.3.1 Gibbs entropy and Boltzmann entropy compared

It can be shown that the standard deviation of energy

$$\Delta E = \sqrt{\langle E^2 \rangle - \langle E \rangle^2} \tag{8.21}$$

yielded by a canonical distribution will, for systems of very many degrees of freedom, be small compared to the expectation value of energy,

$$\frac{\Delta E}{\langle E \rangle} \sim \frac{1}{\sqrt{N}}. \tag{8.22}$$

Recall that, for macroscopic systems, N is on the order of Avogadro's number, that is, on the order of 10^{23}, so the deviation in energy is very small indeed. The energy is almost certain to depart only negligibly from its expectation value, and so the canonical distribution can be replaced, for the purpose of calculating expectation values of thermodynamic quantities, with a micro-canonical distribution on the energy surface corresponding to the expectation value of energy.

Moreover, most of this energy surface will be occupied by the equilibrium macrostate, and there is little difference between calculating the phase-space volume of the energy surface and the volume of its largest macrostate. Thus, for systems in equilibrium and macroscopically many degrees of freedom, the Boltzmann entropy and the Gibbs entropy will be approximately equal, up to a constant, and, crucially, will exhibit the same dependence on on external parameters.

Suppose we extend the identification of (8.15) as entropy for systems other than those in thermal contact with a heat bath. We might even extend this identification to non-equilibrium situations, for which thermodynamic entropy is undefined. Then, because of the measure-preserving property of Hamiltonian flow on phase space, for an isolated system, S_G will not increase with time. This makes it a poor candidate for tracking entropy changes in a process of relaxation to equilibrium. However, we can also define a coarse-grained entropy by partitioning the phase space Γ into small regions of equal volume, and replacing the probability distribution over microstates by one that is uniform over elements of the partition. The idea is that, if the elements of the partition are smaller than our ability to discriminate between microstates, this smeared probability distribution will yield virtually the same probabilities for outcomes of feasible measurements as the fine-grained distribution. The Gibbs entropy associated with this smeared probability distribution can increase with time, capturing the idea that information we have about the earlier state of the system becomes less relevant to outcomes of future measurements (and hence less valuable for our efforts to exploit the system's energy to do work).

Recall that the definition of the Boltzmann entropy also requires a coarse-graining of the phase space of the system. The conceptual differences between Boltzmann entropy and coarse-grained Gibbs entropy are not great.

References

Abrams, M. (2012). Mechanistic probability. *Synthese 187*, 343–375.

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

Albrecht, A. and D. Phillips (2012). Origin of probabilities and their application to the multiverse. `arXiv:1212.0953[gr-qc]`.

Albrecht, A. and L. Sorbo (2004). Can the universe afford inflation? *Physical Review D 70*, 063528.

Allis, W. P. and M. A. Herlin (1952). *Thermodynamics and Statistical Mechanics*. McGraw-Hill Book Company.

Angelopoulos, A., A. Apostolakis, and E. Aslanides (1998). First direct observation of time-reversal non-invariance in the neutral-kaon system. *Physics Letters B 444*, 43–51.

Arntzenius, F. (2004). Time reversal operations, representations of the Lorentz group, and the direction of time. *Studies in History and Philosophy of Modern Physics 35*, 31–43.

Arntzenius, F. and H. Greaves (2007). Time reversal in classical electromagnetism. PhilSci Archive, http://philsci-archive.pitt.edu/archive/00003280/.

Atkins, P. (2007). *Four Laws That Drive the Universe*. Oxford: Oxford University Press.

Barrett, J. A. (1999). *The Quantum Mechanics of Minds and Worlds*. Oxford University Press.

Bassi, A. and G. C. Ghirardi (2003). Dynamical reduction models. *Physics Reports 379*, 257–426.

Batterman, R. W. (2010). Reduction and renormalization. In G. Ernst and A. Hüttemann (Eds.), *Time, Chance, and Reduction: Philosophical Aspects of Statistical Mechanics*. Cambridge: Cambridge University Press. Preprint, differing only in inessentials from the published version, available at `http://philsci-archive.pitt.edu/2852/`.

Beisbart, C. and S. Hartmann (Eds.) (2011). *Probabilities in Physics*. Oxford: Oxford University Press.

Berkovitz, J., R. Frigg, and F. Kronz (2006). The ergodic hierarchy, randomness and Hamiltonian chaos. *Studies in History and Philosophy of Modern Physics 37*, 661–691.

Boltzmann, L. (1877). Bemerkungen über einige Probleme der mechanische Wärmetheorie. *Sitzungberichte der Kaiserlichen Akademie der Wissenschaften. Mathematisch-Naturwissenschaftliche Classe 75*, 62–100. Reprinted in Boltzmann (1909, 116–122).

Boltzmann, L. (1878). Über die Beziehung der Diffusionsphänomene zum zweiten Hauptsatze der mechanischen Wärmetheorie. *Die Sitzungsberichte der Akademie der Wissenschaften, Wien. Mathematisch-Naturwissenschaften Klasse 78*, 733–763.

Boltzmann, L. (1895). On certain questions of the theory of gases. *Nature 51*, 413–415.

Boltzmann, L. ([1896, 1898] 1995). *Lectures on Gas Theory*. New York: Dover Publications.

Boltzmann, L. (1909). *Wissenschaftliche Abhandlung*. Leipzig: J. A. Barth.

Bricmont, J., D. Dürr, M. Galavotti, G. Ghirardi, F. Petruccione, and N. Zanghì (Eds.) (2001). *Chance in Physics*. Number 574 in Lecture Notes in Physics. Berlin: Springer.

Brown, H. R., W. Myrvold, and J. Uffink (2009). Boltzmann's H-theorem, its discontents, and the birth of statistical mechanics. *Studies in History and Philosophy of Modern Physics 40*, 174–191.

Brown, H. R. and J. Uffink (2001). The origins of time-asymmetry in thermodynamics: The minus first law. *Studies in History and Philosophy of Modern Physics 32*, 525–538.

Callender, C. (2000). Is time 'handed' in a quantum world? *Proceedings of the Aristotelian Society 100*, 246–269.

Daub, E. E. (1969). Probability and thermodynamics: The reduction of the second law. *Isis 60*, 318–330.

Dorfman, J. (1999). *An Introduction to Chaos in Nonequlibrium Statistical Mechanics*. Cambridge University Press.

Dugdale, J. S. (1996). *Entropy and its Physical Meaning*. London: Taylor & Francis.

Earman, J. (2002). What time reversal invariance is and why it matters. *International Studies in the Philosophy of Science 16*, 245–264.

Earman, J. (2006). The "Past Hypothesis": not even false. *Studies in History and Philosophy of Modern Physics 37*, 399–430.

Earman, J. and J. Norton (1998). Exorcist XIV: The wrath of Maxwell's Demon. Part I. From Maxwell to Szilard. *Studies in History and Philosophy of Modern Physics 29*, 435–471.

Earman, J. and J. Norton (1999). Exorcist XIV: The wrath of Maxwell's Demon. Part II. From Szilard to Landauer and beyond. *Studies in History and Philosophy of Modern Physics 30*, 1–40.

Eddington, A. S. (1931). The end of the world: from the standpoint of mathematical physics. *Nature 127*, 447–453.

Ehrenfest, P. and T. Ehrenfest ([1912] 1990). *The Conceptual Foundations of the Statistical Approach in Mechanics*. New York: Dover Publications.

Einstein, A. (1954). What is the theory of relativity? In *Ideas and Opinions*, pp. 222–227. New York: Dell.

Engel, E. M. (1992). *A Road to Randomess in Physical Systems*. Berlin: Springer-Verlag.

Garber, E., S. G. Brush, and C. W. F. Everitt (Eds.) (1995). *Maxwell on Heat and Statistical Mechanics: On "Avoiding All Personal Enquiries" of Molecules*. Bethlehem, Pa: Lehigh University Press.

Ghirardi, G. C. (2011). Collapse theories. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2011 ed.).

Gibbs, J. W. (1875). On the equilibrium of heterogeneous substances. *Transactions of the Connecticut Academy of Arts and Sciences 3*, 108–248, 343–524. Reprinted in Gibbs (1961, pp. 55-353).

Gibbs, J. W. (1902). *Elementary Principles in Statistical Mechanics: Developed with Especial Reference to the Rational Foundation of Thermodynamics*. New York: Charles Scribner's Sons.

Gibbs, J. W. ([1906] 1961). *The Scientific Papers of J. Willard Gibbs*. New York: Dover Publications, Inc.

Goldstein, S. (2001). Boltzmann's approach to statistical mechanics. See Bricmont et al. (2001), pp. 39–54. Available online at http://www.math.rutgers.edu/ oldstein/.

Greaves, H. and D. Wallace (2006). Justifying conditionalization: Conditionalization maximizes expected epistemic utility. *Mind 115*, 607–632.

Hacking, I. (1971). Equipossibility theories of probability. *British Journal for the Philosophy of Science 22*, 339–355.

Hacking, I. (1975). *The Emergence of Probability*. Cambridge: Cambridge University Press.

Hjalmars, S. (1977). Evidence for Boltzmann's H as a capital eta. *American Journal of Physics 45*, 214–215.

Hopf, E. (1934). On causality, statistics, and probability. *Journal of Mathematics and Physics 13*, 51–102.

Hopf, E. (1936). Über die Bedeutung der willkürlichen Funktionen für die Wahrscheinlichkeitstheorie. *Jahresbericht des Deutschen Mathematiker-Vereinigung 46*, 179–195.

Huang, K. (1986). *Statistical Mechanics* (Second ed.). John Wiley & Sons.

Jackson, E. A. (1968). *Equilibrium Statistical Mechanics*. New York: Dover Publications.

Khinchin, A. I. (1949). *Mathematical Foundations of Statistical Mechanics*. New York: Dover Publications.

Knott, C. G. (1911). *Life and Scientific Work of Peter Guthrie Tait*. Cambridge: Cambridge University Press.

Leeds, S. (2006). Discussion: Malament on time reversal. *Philosophy of Science 73*, 448–458.

Lees, J. P., V. Poireau, V. Tisserand, J. G. Tico, E. Grauges, A. Palano, G. Eigen, B. Stugu, and D. N. Brown (2012). Observation of time-reversal violation in the B^0 meson system. *Physical Review Letters 109*, 211801.

Leff, H. S. and A. F. Rex (Eds.) (2003). *Maxwell's Demon 2: Entropy, Classical and Quantum Information, Computing*. Bristol and Philadelphia: Institute of Physics Publishing.

Lewis, D. (1980). A subjectivist's guide to objective chance. In R. C. Jeffrey (Ed.), *Studies in Inductive Logic and Probability*, Volume II, pp. 263–93. University of Caifornia Press.

Malament, D. B. (2004). On the time reversal invariance of classical electromagnetic theory. *Studies in History and Philosophy of Modern Physics 35*, 295–315.

Malament, D. B. and S. L. Zabell (1980). Why Gibbs phase averages work—the role of ergodic theory. *Philosophy of Science 47*, 339–349.

Maxwell, J. C. (1871). *Theory of Heat*. London: Longmans, Green, and Co.

Maxwell, J. C. ([1872] 2001). *Theory of Heat*. New York: Dover Publications.

Maxwell, J. C. (1878a). Diffusion. In *Encyclopedia Britannica* (Ninth ed.), Volume 7, pp. 214–221. Reprinted in Niven (1965, pp.625–646).

Maxwell, J. C. (1878b). Tait's "Thermodynamics". *Nature 17*, 257–259, 278–280. Reprinted in Niven (1965, pp. 660-671).

Maxwell, J. C. ([1891] 1954). *A Treatise on Electricity & Magnetism*. New York: Dover Publications.

Myrvold, W. C. (2011). Statistical mechanics and thermodynamics: A Maxwellian view. *Studies in History and Philosophy of Modern Physics 42*, 237–243.

Myrvold, W. C. (2012a). Determistic laws and epistemic chances. In Y. Ben-Menahem and M. Hemmo (Eds.), *Probability in Physics*, pp. 73–85. Springer.

Myrvold, W. C. (2012b). Probabilities in statistical mechanics: What are they? Available at http://philsci-archive.pitt.edu/9236/.

Myrvold, W. C. (2012c). Probabilities in statistical mechanics: What are they? Available at http://philsci-archive.pitt.edu/9236/.

Newton, I. ([1730] 1952). *Opticks*. New York: Dover Publications.

Niven, W. D. (Ed.) (1965). *The Scientific Papers of James Clerk Maxwell*, Volume Two. New York: Dover Publications. Reprint of Cambridge University Press edition of 1890.

Norton, J. (2005). Eaters of the lotus: Landauer's principle and the return of Maxwell's demon. *Studies in History and Philosophy of Modern Physics 36*, 375–411.

Norton, J. (2011). Waiting for Landauer. Available at http://philsci-archive.pitt.edu/8635/.

Pauli, W. (1973). *Pauli Lectures on Physics: Volume 3. Thermodynamics and the Kinetic Theory of Gases*. Cambridge, MA: The MIT Press.

Penrose, O. (2001). The direction of time. See Bricmont et al. (2001), pp. 61–82. Available online at http://www.ma.hw.ac.uk/ oliver/.

Penrose, R. (1990). *The Emperor's New Mind*. Oxford: Oxford University Press.

Poincaré, H. (1912). *Calcul des probabilités* (2nd ed.). Paris: Gauthier-Villars.

Price, H. (1996). *Time's Arrow and Archimedes' Point*. Oxford: Oxford University Press.

Price, H. (2006). The thermodynamic arrow: puzzles and pseudo-puzzles. In I. Bigi and M. Faessler (Eds.), *Time and Matter: Proceedings of the International Colloquium on the Science of Time*, pp. 209–224. Singapore: World Scientific.

Rees, M. (1997). *Before the Beginning: Our Universe and Others*. Addison-Wesley.

Rosenthal, J. (2010). The natural-range conception of probability. In G. Ernst and A. Hüttemann (Eds.), *Time, Chance and Reduction: Philosophical Aspects of Statistical Mechanics*, pp. 71–91. Cambridge: Cambridge University Press.

Rosenthal, J. (2012). Probabilities as ratios of ranges in initial-state spaces. *Journal of Logic, Language, and Information 21*, 217–236.

Saunders, S. (2006). On the explanation for quantum statistics. *Studies in History and Philosophy of Modern Physics 37*, 192–211.

Savage, L. J. (1973). Probability in science: A personalistic account. In P. Suppes (Ed.), *Logic Methodology, and Philosophy of Science IV*, pp. 417–428. Amsterdam: North-Holland.

Sklar, L. (1993). *Physics and Chance*. Cambridge University Press.

Strevens, M. (2003). *Bigger than Chaos: Understanding Complexity through Probability*. Cambridge, MA: Harvard University Press.

Strevens, M. (2011). Probability out of determinism. In Beisbart and Hartmann (2011, 339–364).

Thomson, W. (1874). Kinetic theory of the dissipation of energy. *Nature 9*, 441–444.

Tolman, R. C. (1938). *The Principles of Statistical Mechanics*. Oxford: Clarendon Press. Reprint, Dover Publications, 1979.

Uffink, J. (2001). Bluff your way in the second law of thermodynamics. *Studies in History and Philosophy of Modern Physics 32*, 305–394.

Uffink, J. (2007a). Compendium of the foundations of classical statistical physics. In J. Butterfield and J. Earman (Eds.), *Handbook of the Philosophy of Science. Philosophy of Physics, Part B*, pp. 923–1074. Amsterdam: North Holland. Available online at http://philsci-archive.pitt.edu/archive/00002691.

Uffink, J. (2007b). Compendium of the foundations of statistical physics. In J. Butterfield and J. Earman (Eds.), *Handbook of the Philosophy of Science: Philosophy of Physics*, pp. 924–1074. Amsterdam: North-Holland.

van Fraassen, B. (1989). *Laws and Symmetry*. Oxford: Oxford University Press.

von Kries, J. (1886). *Die Principien Der Wahrscheinlichkeitsrechnung: Eine Logische Untersuchung*. Frieburg: Mohr.

von Plato, J. (1983). The method of arbitrary functions. *The British Journal for the Philosophy of Science 34*, 37–47.

Weinberg, S. (1995). *The Quantum Theory of Fields*, Volume 1. Cambridge University Press.

Winsberg, E. (2004). Can conditioning on the "Past Hypothesis" militate against the reversibility objections? *Philosophy of Science 71*, 489–504.