Typicality, Irreversibility and the Status of Macroscopic Laws

Dustin Lazarovici*, Paula Reichert†

Ludwig-Maximilians-Universität München
Mathematisches Institut

Abstract

We discuss Boltzmann's probabilistic explanation of the second law of thermodynamics, providing a comprehensive presentation of what is called today the *typicality account*. Countering its misconception as an alternative explanation, we examine the relation between Boltzmann's H-theorem and the general typicality argument, demonstrating the conceptual continuity between the two. We then discuss the philosophical dimensions of the concept of typicality and its relevance for scientific reasoning in general, in particular for understanding the reduction of macroscopic laws to microscopic laws. Finally, we reply to recent criticisms of the typicality account.

*lazarovici@math.lmu.de
†reichert@math.lmu.de

1 Introduction

Over the last two decades, a series of papers by various distinguished mathematical physicists stressed the importance of the concept of *typicality* as a basis for probabilistic reasoning in physics, in particular as a basis for the explanation of the second law of thermodynamics in statistical mechanics (Lebowitz, 1993; Bricmont, 1995; Penrose, 1999; Goldstein, 2001). None of the authors took much credit for the ideas he presented, each of them rather stressed that he was recapturing or reformulating the groundbreaking insights of Ludwig Boltzmann who, more than one century ago, had shown how to explain and derive macroscopic regularities from the underlying laws governing the motion of the microscopic constituents of matter. However, reintroducing these ideas to physicists, mathematicians and philosophers proved to be highly necessary as their relevance is rarely appreciated today and the response to the papers of Lebowitz, Goldstein and others shows that they are still subject to widespread misconceptions and misunderstandings.

In this paper, we provide a comprehensive summary of the typicality account, spelling out some of the details and subtleties that have remained unspoken in the before-mentioned presentations, thus having left room for objections and misunderstandings that we hope to eliminate (Section 2). We also demonstrate the conceptual continuity between the H-theorem and the general typicality account, showing that it is false that they are often viewed as alternatives (Section 3). Putting things in wider perspective, we discuss the relevance of typicality for scientific reasoning in general, in particular for understanding the reduction of macroscopic laws to microscopic laws (Section 4). Finally, we address the most common objections against the typicality account that have been raised in the contemporary literature (Section5).

2 Boltzmann's statistical mechanics

2.1 The typicality account

Our discussion is concerned with the explanation of the irreversible thermodynamic behavior of macroscopic systems. The term "thermodynamic behavior" thereby refers to the ubiquitous phenomenon that physical systems, prepared or created in a non-equilibrium state and then suitably isolated from the environment, tend to evolve to and then stay in a distinguished macroscopic configuration called the *equilibrium state.* Familiar examples are the spreading of a gas, the mixing of milk and coffee, the disappearance of temperature gradients, and so on.

Historically, this empirical regularity was captured by the *second law of thermodynamics*, positing the monotonous increase of a macroscopic variable of state called *entropy*, which attains its maximum value in equilibrium. The main task of *statistical mechanics* is to explain this macroscopic regularity on the basis of the underlying laws guiding the behavior of the system's micro-constituents.

A crucial ingredient to the understanding of this issue is the distinction between macro- and microstate of a system. Whereas the microstate $X(t)$ of a system is given by the complete specification of all its microscopic degrees of freedom, its macrostate $M(t)$ is specified in terms of physical variables that characterize the system on macroscopic scales (like it's volume, pressure, temperature, and so on). The macroscopic state of a system is completely determined by its microscopic configuration, that is $M(t) = M(X(t))$, but one and the same macrostate can be realized by a large number of different microstates all of which "look macroscopically the same". The partitioning of the set of microstates into sets corresponding to macroscopically distinct states is therefore called a "coarse-graining". Turning to the phase-space picture of Hamiltonian mechanics for an N-particle system, a microstate corresponds to one point $X = (q, p)$ in phase-space $\Omega \cong \mathbb{R}^{3N} \times \mathbb{R}^{3N}$, $q = (q_1, q_2, ..., q_N)$ being the position- and $p = (p_1, p_2, ..., p_N)$ the momentum-coordinates of the N particles, whereas a macrostate M corresponds to an entire region $\Gamma_M \subseteq \Omega$ of phase-space, namely the set of all microstates that realize M. The microscopic laws of motion are such that any initial microstate X_0 determines the complete microevolution $X(t) = \phi_t(X_0)$ of the system, represented by a unique trajectory in phase-space going through X_0, thereby also determining its complete macro-evolution $M(X(t))$ as the microstate passes through different macro-regions.

These concepts are pretty much forced on us if we accept the supervenience of macroscopic facts on microscopic facts and they are essential to understanding the nature of the problem. The second law of thermodynamics describes an empirical regularity about the *macro-evolution* $M(t)$ of a physical system. This macro-evolution, however, supervenes on the evolution of the system's microscopic configuration which is determined by precise and unambiguous laws of motion. The aspiration of statistical mechanics is thus to explain or justify the empirical regularity expressed in the macroscopic law on the basis of the underlying microscopic theory. This seems like a quite formidable task, though, as it requires us to reconcile the *irre-*

versibility of thermodynamic behavior with the *time-reversal symmetry* of the microscopic laws of motion. The task was nevertheless accomplished by Ludwig Boltzmann at the end of the 19th century and we recall that his account was crucially based on two profound insights:

1. The identification of the (Clausius) entropy with the (logarithm of) the phase-space volume corresponding to its current macrostate. Formally:

$$S = k_B \ln |\Gamma_{M(X)}|, \tag{1}$$

   where k_B is the Boltzmann constant and $|\Gamma_M|$ denotes the volume (the Lebesgue or Liouville measure) of the phase-space region Γ_M. The Boltzmann entropy is thus de facto a logarithmic measure of the phase-space volume corresponding to the system's macrostate.

2. The understanding that the equilibrium macro-region occupies almost the entire phase-space volume, i.e., that almost every microstate is an equilibrium state. Note that the logarithm in the definition of the Boltzmann entropy has the effect that significant differences in entropy correspond to huge differences in the phase-space volume corresponding to the respective macrostates. And indeed, we will generally find that for macroscopic systems, i.e. for systems with a very large number of microscopic degrees of freedom, the partitioning of microstates into macrostates does not correspond to a partitioning of phase-space into regions of roughly the same size, but into regions whose sizes vary by a great many orders of magnitude, with the region of maximum entropy – by definition the equilibrium region – being by far the largest.

These two insights are the key ingredients in Boltzmann's account of the second law of thermodynamics. What we learn from them is, first and foremost, that the thermodynamic behavior that we want to explain is not a feature of certain *particular* micro-evolutions, but rather the kind of macro-behavior that would correspond to almost any generic trajectory that the configuration of a macroscopic system, starting in a non-equilibrium region, could follow through phase-space. Indeed, the dynamics of a system whose microscopic configuration starts out in a tiny non-equilibrium macro-region would have to be extremely peculiar to *avoid* carrying the micro-state into larger and larger phase-space volumes, corresponding to gradually increasing entropy, and finally into the equilibrium region where it will spend by far most of the time (except for small fluctuations of the entropy about its maximum value). This is why accounts of the second law insisting that the explanation of thermodynamic behavior must be grounded in some special property pertaining to solutions of the microscopic equations of motion, e.g., they being ergodic (Frigg and Werndl, 2011, 2012), are on the wrong track, disregarding the extraordinary robustness of Boltzmann's arguments against the details of the microscopic theory.

Nevertheless, critics have devoted multiple publications and many pages to pointing out that, if we consider a system in a initial non-equilibrium microstate, we cannot conclude that it *must* evolve into equilibrium solely on the basis that the equilibrium-region is vastly larger than the non-equilibrium region, covering almost the entire phase-space (Frigg, 2009, 2011, see also the discussion in Section 5). And while this is, as such, a correct observation, it is crucial to understand that this is *not* what the typicality account, in the end, claims (and it is not what Boltzmann had claimed, at least since 1876). Indeed, we know for a fact that, given a low-entropy macro-state M_2, there exist initial conditions in the corresponding macro-region Γ_{M_2} that will *not* evolve to equilibrium but follow a trajectory of *decreasing* entropy instead. And this is not, in the first place, due to any involved mathematical or philosophical argument, but a straightforward consequence of the *reversibility* of the microscopic laws as was famously pointed out by Johann Loschmidt in 1876. So Lebowitz rightly warned us, quoting Ruelle, that the ideas of Boltzmann are "at the same time simple and rather subtle" (Lebowitz, 1993, p. 7).

To discuss these subtleties and make the typicality reasoning more precise, we turn to the paradigmatic example of a gas in a box. We thus consider a system of about $N = 10^{23}$ particles, interacting by a repelling short-range potential, which are confined to a finite volume within a box with reflecting walls. Now assume that we find or prepare the system in the macrostate M_2 sketched below (Fig. 1), that is, we consider a particle configuration that looks, macroscopically, like a gas filling out half the volume. What kind of evolution, on the macroscopic scale, should we expect for the gas?
Well, a simple combinatorial argument, given by Boltzmann, shows that *the overwhelming majority* of microstates that the system could possibly evolve in will look, macroscopically, like M_{eq}, i.e. like a gas that is homogeneously distributed over the entire volume of the box. In fact, one can readily conclude that the phase-space volume corresponding to this *equilibrium marostate* M_{eq} is about $2^N \approx 10^{10^{23}}$ times (!) larger than the phase-space

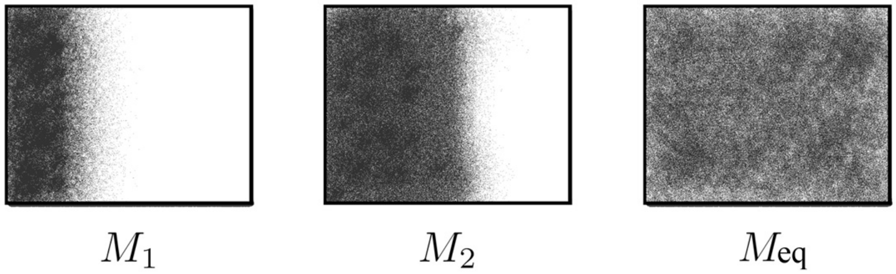

Figure 1: Thermodynamic evolution of a gas

volume occupied by configurations with substantially lower entropy. Hence, as the particles move with different speeds in different directions, scattering from each other and occasionally from the walls, the system's microstate wanders around on an erratic path in the high-dimensional phase-space and we should expect, by all reasonable means, that this path will soon end up in the equilibrium region $\Gamma_{M_{eq}}$ (and then leave $\Gamma_{M_{eq}}$ only very rarely, corresponding to small fluctuations of the entropy about its maximal value).[1] However, it is clear (and it was clear to Boltzmann) that there are initial conditions in Γ_{M_2} for which the system will not exhibit this "expected" (we will later say *typical*) macro-behavior, but follow an anti-thermodynamic trajectory. For if we consider a macrostate of even lower entropy, M_1, the time-reversal symmetry of the equations of motions implies that for every solution corresponding to a macro-evolution from M_1 to M_2, there exists another solution carrying an initial microstate in Γ_{M_2} into the lower-entropy macro-region Γ_{M_1}. And yet, as Boltzmann understood, the microstates (the initial conditions in Γ_{M_2}) that lead to such an anti-thermodynamic evolution are *extremely special ones* relative to all possible microstates realizing M_2. The correct statement is thus that *almost all* initial microstates in Γ_{M_2} will evolve into the equilibrium-region Γ_{eq}, while only a very small set of "bad" initial conditions will show the anti-thermodynamic evolution from Γ_{M_2} into Γ_{M_1}. We will make these arguments more precise in a minute.

For now, let us emphasize that it's more appropriate not to consider any

[1] Larger fluctuations, i.e. from M_{eq} back into M_2, are *possible* as well. However, as (Boltzmann, 1896b) already noted, the time-scales on which that substantial fluctuations are to be expected are so astronomical – about $10^{10^{20}}$ years for the gas model – that they have absolutely no empirical relevance.

individual trajectory, but the set of all solutions with initial condition in Γ_{M_2}. The dynamics of a system of about $N \sim 10^{23}$ particles are highly chaotic, in the sense that even the slightest variation in the initial configuration can lead to considerable differences in the time-evolution. Under the Hamiltonian dynamics, the set of microstates realizing M_2 at the initial time will thus quickly spread all over phase-space (respectively a hypersurface compatible with its constants of motion) with the overwhelming majority of microstates ending up in the equilibrium-region and only a small fraction of special initial configurations evolving into the comparably tiny macro-regions of equal or lower entropy.

All in all, Boltzmann's analysis tells us that it cannot be true that *every* non-equilibrium configuration will follow the second law of thermodynamics and undergo an evolution of increasing entropy. We can, however, assert that *typical* microscopic configurations, realizing a low-entropy initial macrostate, will evolve into equilibrium and stay in equilibrium for most of the time.

2.2 The measure of typicality

Throughout this argument, the intuitive notions of *almost all* and *extremely special*, that we used synonymously to *typical/atypical*, were understood in terms of the stationary *Liouville measure*, i.e. in terms of the *phase-space volume* of the set of microstates with the relevant property. More precisely, for a perfectly isolated system with total energy E, we would have to consider not the Liouville measure but the induced *microcanonical measure* μ_E on the hypersurface $\Gamma_E \subset \Omega$, to which the motion of the system is confined in virtue of the energy conservation. For simplicity, we will omit this distinction and merely refer to "phase-space" and the "measure" or "size" of macro-regions.

In any case, a crucial property of the Liouville measure as well as the microcanonical measure is their *stationarity* under the microscopic time-evolution. Intuitively, this means that the Hamiltonian flow ϕ_t behaves like an incompressible fluid on phase-space. Formally, it means that for all measurable sets $A \subseteq \Omega$ and all times $t \in \mathbb{R}$, we have $|\phi_t(A)| = |A|$. This is such an essential feature because it means that

a) the notion of typicality is *timeless*, i.e. a typicality statement does not depend on a reference to any external time-parameter.

b) the Hamiltonian dynamics "care about" the measure of the macro-regions that play such a central role in the argument in the sense that the stationary measure as a measure on *initial conditions* carries over to a well-defined measure on *solution trajectories*, which is such that the number of trajectories passing through a phase-space region at any given time is proportional to the size of that region.

Turning back to Boltzmann's explanation of the second law, we note that the Liouville measure (respectively the microcanonical measure) as a *typicality measure* serves two purposes in the argument:

1. To establish that the region of phase-space corresponding to the macrostate M_2 is very much larger than the region of phase-space corresponding to the macrostate M_1, and that the region of phase-space corresponding to the equilibrium macrostate M_{eq} is very much larger than the region of phase-space corresponding to the macrostate M_2, so large, in fact, that it occupies almost the entire phase-space volume.

   It is easy to learn about this "dominance of the equilibrium state" (Frigg, 2009) and yet hard to appreciate what it is really saying, since the scale of the proportions expressed by the innocuous term "almost entirely" are beyond anything that we could intuitively grasp (just think of the ratio $10^{10^{23}} : 1$ for the gas-model).

2. To define a notion of typicality *relative to the current macrostate of the system*, allowing us to assert, for instance, that almost all microstates in the non-equilibrium region Γ_{M_2} will evolve into equilibrium.

   Regarding the meaning of "almost all", one should note that it's only in the idealized situation of a *thermodynamic limit* (where the number of microscopic degrees of freedom goes to infinity) that one can expect the exception set of "bad" configurations to be of measure *zero*, while if we argue about a realistic system, the *atypicality* of such configurations is substantiated by the fact that they have very very small (though positive) measure compared to that of all microstates realizing M_2.

   In fact, stationarity of the Liouville measure allows us to estimate the measure of the good microstates relative to the bad microstates in Γ_{M_2} by the ratio of phase-space volume occupied by M_2 to the phase-space volume corresponding to states of lower entropy. For let $B \subset \Gamma_{M_2}$ be the set of initial conditions that will have evolved into a lower-entropy region Γ_{M_1} after a time Δt, then $\Phi_{\Delta t}(B) \subseteq \Gamma_{M_1}$ and thus $|B| = |\Phi_{\Delta t}(B)| \leq |\Gamma_{M_1}|$, so that $|B| : |\Gamma_{M_2}| \approx |\Gamma_{M_1}| : |\Gamma_{M_2}| \approx 1 : 10^{10^{23}}$.

2.3 Irreversibility

By incorporating into our analysis what is essentially Boltzmann's answer to Loschmidt's reversibility objection, we have already seen the solution to the problem that seemed like the greatest challenge to our reductionist enterprise: the *prima facie* contradiction between the irreversibility of thermodynamic processes and the reversibility of the underlying mechanical laws. To emphasize how this apparent contradiction is resolved, we recall that it was essential to our argument that it always referred to (typical or atypical) initial conditions *relative to the initial macrostate.* Of course, in terms of overall phase-space volume, a non-equilibrium macrostate occupies a vanishingly small fraction of phase-space to begin with, corresponding (if you will) to a very low a priori probability. The relevant notion of typicality when discussing convergence to equilibrium from a non-equilibrium macrostate M_2 is thus defined by the Liouville measure conditioned on the fact that the initial microstate is in the respective phase-space region Γ_{M_2}.

Now, as we already observed, the time-symmetry of the microscopic laws manifests itself in the fact that the phase-space volume occupied by the bad initial conditions in Γ_{eq}, for which the system will fluctuate out of equilibrium into the macrostate M_2 (let's say), is just as large as the phase-space volume occupied by the good initial conditions in Γ_{M_2} for which the system will relax into equilibrium. In other words, over any given period of time, there are just as many solutions that evolve *into* equilibrium, as there are solutions evolving *out* of equilibrium into a lower entropy state, but the first case is nevertheless *typical* for systems in non-equilibrium, whereas the second case is *atypical* with respect to all possible equilibrium configurations in Γ_{eq}. It is this fact and this fact alone that establishes the *irreversibility* of thermodynamic behavior.

2.4 The Past Hypothesis and the thermodynamic arrow

By telling us that the origin of the thermodynamic asymmetry that is expressed in the 'second law' lies only in the specialness of the initial low-entropy macrostates, the typicality account is shifting the explanatory burden from why it is that a system in non-equilibrium typically relaxes to equilibrium (once macroscopic constraints are removed), to why it is that we find systems in such special states *in the first place.* Note that a typical configuration *simpliciter*, i.e. a typical configuration with respect to *all possible microstates*, is a state for which the system is in equilibrium, will be in equilibrium for most of its future and has been in equilibrium for most of its past – which thus describes a time-symmetric situation.

Of course, as long as we are preoccupied with boxes of gas, or melting ice-cubes, or other confined systems, their low-entropy states will always be attributable to influences from outside, i.e. to the fact that these systems are actually part of some larger system (usually containing a physicist, or a freezer, or the like) before "branching off" to undergo a more or less autonomous evolution as more or less isolated subsystems. This presupposes, however, that those larger systems have been out of equilibrium *themselves*, otherwise they could not have given rise to subsystems with less then maximal entropy without violating the second law. And if we think this through to the end, we finally arrive at the question why it is that we find *our universe* in such a special state, far away from equilibrium (and how we justify our believe that it was even further away from equilibrium the farther we go back in the past). This is what Goldstein calls the "hard part of the problem [of irreversibility]" (Goldstein, 2001, p. 49) and it concerns, broadly speaking, the *origin* of irreversibility and the thermodynamic arrow of time in our universe. Dealing with the "hard part" would require us to discuss the meaning and the status of the *Past Hypothesis*[2] stipulating a very-low-entropy initial state of our universe. However, this issue is beyond the scope of the present paper and we shall return to it only briefly in the course of our discussion.

3 Typicality and the H-theorem

Although the formula engraved on Boltzmann's tombstone is equation (1), connecting the entropy of a microstate with the "probability" of the corresponding macrostate, his name is at least as intimately associated with the Boltzmann equation and the H-theorem, describing, in a more quantitative manner, convergence to equilibrium for a low-density gas. This H-theorem is of great interest in the light of our previous discussion, first, because it illustrates very clearly the need for a typicality argument and second, because it can be viewed as a concrete implementation of the general scheme that we've just presented. In this context, we want to counter a common misconception that has most likely arisen from Boltzmann's first presentation of

[2]The term "Past Hypothesis" is due to (Albert, 2000), though the necessity of such an assumption was already noted by Boltzmann (1896a, pp. 252-253). See also (Feynman, 1967) and (Carroll, 2010) for a very nice discussion.

the H-theorem and persisted despite his more refined argumentation in later writings, namely that the H-theorem and the typicality account are somehow *competing* accounts of macroscopic irreversibility and the convergence to equilibrium. Huw Price, for instance, writes with respect to the latter:

> *In essence, I think – although he himself does not present it in these terms – what Boltzmann offers is an alternative to his own famous H-Theorem. The H-theorem offers a dynamical argument that the entropy of a non-equilibrium system must increase over time, as a result of collisions between its constituent particles. [...] The statistical approach does away with this dynamical argument altogether.* (Price, 2002, p. 27)

The pertinent entry in the Stanford Encyclopedia of Philosophy (Uffink, 2008), too presents Boltzmann's work as a series of rather incoherent (and ultimately wanting) attempts to explain the second law.

We are convinced that the reason why Boltzmann did not present the "statistical approach" as an alternative to the H-theorem is that, in fact, it isn't. Understood correctly, there is a clear conceptual continuity between the H-theorem and the typicality account so that the latter does not appear as a break with Boltzmann's earlier work, but as a distillation of its essence (cf. Goldstein, 2012; Goldstein and Lebowitz, 2004). Understanding this connection, we will also see that the main objection raised against the conclusiveness of the H-theorem, concerning its account of thermodynamic irreversibility, is unfounded. To make this case, we shall first review what the H-theorem is about and how it's grounded in the microscopic theory.[3]

3.1 The H-theorem

Recall that the microstate of an N-particle system is represented by a point $X = (q_1, ..., q_N; p_1, ..., p_N) \in \Omega$ in $6N$-dimensional phase-space, comprising the position and momenta of all particles. The same state (modulo permutations of the particles) can also be represented as N points in the 6-dimensional *μ-space*, whose coordinates correspond to position and velocity of a *single particle*, i.e. $X \to \{(q_1, v_1), ..., (q_N, v_N)\}$, with $v_i := p_i/m$. The H-theorem is concerned with the evolution of a function $f_X(q, v)$ on μ-space, that is supposed to provide an efficient description of the most important (macroscopic) characteristics of the gas in the microstate X. This

[3] For a good introduction, see, for instance, (Davies, 1977). For a detailed mathematical treatment, see (Spohn, 1991), (Villani, 2002), and (Lebowitz, 1981).

function is defined as the *empirical distribution* or *coarse-grained density* of points in μ-space. In principle, one can think of dividing μ-space into little cells whose dimension is large enough to contain a great number of particles, yet very small compared to the resolution of macroscopic observations, and counting the number of particles in each cell. For fixed q and v, $f_X(q,v)$ thus corresponds to the proportion of particles located *near* q with velocity *approximately* v. In the limit where the size of the cells goes to zero, the empirical distribution becomes the actual distribution

$$f_{X(t)}(q,v) := \frac{1}{N}\sum_{i=1}^{N}\delta(q-q_i(t))\,\delta(v-\frac{1}{m}p_i(t)).$$

We give this formula to emphasize that, although $f_X(q,v)$ is technically a *probability measure*, there's is nothing *random* about it. In fact, it's more adequate to think of it as a *macroscopic variable*, determined, as it always is, by the microscopic configuration of the system. In particular, the distribution function does not describe a random system or an ensemble of systems, but pertains to a *coarse-grained description* of an *individual* system, so that every microstate X determines a unique $f_X(q,v)$, while many different microscopic configurations will coarse-grain to the same μ-space density.

Now the first crucial result is that although the empirical distribution can be different for different microscopic configurations X, it is in fact (more or less) the same for an *overwhelming majority* of possible X. That is, one can show that for typical $X \in \Gamma$, the distribution function is of the form

$$f_X(q,v) \propto e^{-\frac{1}{2}m\beta v^2},$$

for some constant β that is later identified with the inverse temperature of the system. This is the famous *Maxwell* or *Maxwell-Boltzmann distribution*, which is hence the *equilibrium distribution* of the gas. The distribution having no q-dependence means that the gas is homogeneously distributed over the entire volume with no correlations between position and velocities, i.e. with *uniform temperature.*

The goal of Boltzmann's famous H-theorem is thus to show the convergence of an initial *non-equilibrium* distribution $f_0(q,v)$ to the Maxwell-distribution $f_{eq}(q,v)$. The result is thereby based on three claims:

1) For a low-density gas, the time-evolution of $f_{X(t)}(q,v)$ is well described by an effective equation now known as the *Boltzmann equation.*

Starting with an initial distribution $f_0(q,v) = f_{X_0}(q,v)$, it's important to distinguish the function $f_{X(t)}(q,v)$ – whose time-evolution is always determined by that of the microstate $X(t)$ – from the solution $f(t,q,v)$ of the Boltzmann equation with initial condition $f(0,q,v) = f_0(q,v)$ (respectively a smooth approximation thereof). The relevant claim is then that *for typical initial conditions*, $f_{X(t)}(q,v)$ will be (in a precisely specified way) *close* to $f(t,q,v)$ for a sufficiently long period of time, thus providing an *effective* description of the system's time-evolution.

2) For a solution $f(t,q,v)$ of the Boltzmann-equation, the *H-function*

$$H(f(t,q,v)) := \int f(t,q,v) \log f(t,q,v) \mathrm{dqdv}$$

is monotonously decreasing in t.[4]

3) The H-functional reaches its *minimum* for the Maxwell-distribution $f_{eq}(q,v)$.

Together with 2) this implies, in particular, that the Maxwell-distribution is a *stationary* solution of the Boltzmann-equation.

Statements 2) and 3) are fairly standard mathematical results. The crux of the matter lies in statement 1). When Boltzmann first presented the H-theorem in 1872, he argued that a diluted gas *must* evolve in accord with his equation; he later had to mitigate this statement claiming, in effect, only that it would *typically* do so. Indeed, we will see that 1), and therefore the H-theorem, are genuinely typicality statements.

3.2 The *Stoßzahlansatz*

Boltzmann's derivation of what is now known as the Boltzmann equation is famously based on the *Stoßzahlansatz* or the assumption of *molecular chaos*.[5] This is an assumption about the *relative frequencies* of collisions between the particles in the gas. Denoting by $\mathcal{N}(t,q;v_1,v_2)$ the number of collisions happening near q in a small time-interval around t between particles with velocity (approximately) v_1 and v_2, the Stoßzahlansatz is:

$$\mathcal{N}(t,q\,;v_1,v_2) \propto N^2 f(t,q,v_1) f(t,q,v_2)\,|v_1 - v_2|\, \mathrm{d}t\, \mathrm{d}q\, \mathrm{d}v_1 \mathrm{d}v_2, \qquad (2)$$

[4] While the "true" microscopic $H(f_{X(t)}(q,v))$ fluctuates and only decreases "on average".

[5] *Assumption*, unfortunately, is not a perfectly accurate translation of the German word *Ansatz*. Whereas the first is sometimes used synonymously with a logical *premise*, the later has a distinctly pragmatic element and can refer to something more akin to an "approximation" or a "working hypothesis".

i.e. the relative frequency of scattering events between particles of different velocities happening in the cell around q is assumed to be proportional to the density of particles with the respective velocities near the respective position. The scattering probability being proportional to the product of $f(t,q,v_1)$ and $f(t,q,v_2)$ means that particles of different velocities are assumed to be *statistically independent* as they contribute to collisions. This is, more specifically, the meaning of *molecular chaos*.

Boltzmann's derivation, although a brilliant physical argument, was far from a rigorous proof. There are many mathematical subtleties involved in statement 1), concerning, for instance, the existence and uniqueness of solutions to the Boltzmann equation. However, if we can generously overlook these technical points, it is true that *if* and *as long as* the assumption of molecular chaos and equation (2) are valid, statement 1) is correct. Hence, we have to ask: What is the status of molecular chaos and how is it justified?

First and foremost, we have to keep in mind that there is nothing *random* about the interactions in a gas. Which particles are going to collide and how they are going to collide is completely determined by the initial conditions and the microscopic laws of motion. For the purpose of illustration, let's imagine that we could freeze the system at time $t = 0$ and arrange the position and momentum of every single particle before letting the clock run and the system evolve according to the deterministic laws of Newtonian mechanics.[6] We could then, for instance, arrange the initial state in such a way that "slow" particles will almost exclusively scatter with other "slow" particles and "fast" particles with other "fast" particles. But such initial configurations are, obviously, very special ones. For *typical* microscopic configurations, coarse-graining to the initial distribution $f_0(q,v)$, we will however find that the relative frequency with which particles of different velocities meet for the first collision is roughly proportional to the density of particles with the respective velocities near the respective position, i.e. given by eq. (2). This is nothing more and nothing less than the *law of large numbers*, based, in effect, on simple combinatorics. The validity of (2) *at the initial time* is thus, as all law-of-large-number statements, a *typicality statement* and as such another mathematical fact.

We observe here the fundamental difference between the probability density $f(t,q,v)$ and the typicality measure. The "scattering probability" at time t is defined in terms of $f(t,q,v)$, though it's only for typical ini-

[6] Note that there is no issue here as to whether we let the clock run "forwards" or "backwards", the problem is symmetric with respect to the time-evolution in both directions.

tial conditions that the relative frequency of scatterings is actually close to the expectation value. And typical initial conditions are defined, as usual, by the microcanonical measure restricted to the initial macro-region $\Gamma_0 := \{X \in \Gamma_E \mid f_X(q,v) = f_0(q,v)\}$.

This brings us, finally, to the critical part of the H-theorem. For assume that after an (infinitesimal) time-interval Δt for which the validity of the Boltzmann-equation is established, the distribution function has evolved into $f(\Delta t, q, v)$. How do we know that (2) is still a good approximation for all but a small set of initial conditions? It is still true that eq. (2) is satisfied for typical microscopic configurations realizing the *current* distribution, i.e. counting all possible configurations that coarse-grain to $f(\Delta t, q, v)$. But we cannot count all these configurations, since the microstates relevant to our considerations are constraint by the condition that they have evolved from the macro-region corresponding to the initial distribution $f_0(q,v)$. Mathematically, these dynamical constraints on the "combinatorics" translate into the statement that the μ-space coordinates of the particles at time $t > 0$ are no longer *statistically independent*, making it prima facie questionable whether a law-or-large-number statement for the relative frequencies of particle collisions, i.e. molecular chaos, still holds. This is, notably, the only meaningful way in which interactions *build up correlations* and we note, in particular, that the situation is still identical with respect to the time evolution towards the future as well as towards the past of the distinguished initial state.

Now Boltzmann's *Stoßzahlansatz* can be understood as the assumption that statistical independence is preserved by the microscopic time-evolution, or, in other words, that the relative frequency of collisions is always the typical one with respect to the current empirical distribution ($\approx$ the current macrostate). The mathematician refers to such a proposition as *propagation of molecular chaos*. Deriving the Boltzmann equation from a microscopic model, in a rigorous mathematical sense, is thus to validate this *ansatz*, i.e. to show that for typical initial conditions equation (2) remains *approximately* satisfied on sufficiently long time-scales. Sufficiently long, that is, to describe the thermodynamic evolution of a gas into equilibrium.

So, does molecular chaos propagate? That is, do the dynamics of a gas preserve statistical independence well enough to justify the *Stoßzahlansatz*? Based on physical intuition and various encouraging results, there is no reasonable doubt that the answer is affirmative. Given the fact that the microscopic dynamics are highly chaotic, that the number of particles in a gas is huge and that the gas, by assumption, very diluted so that re-collisions (collisions between particles that have already collided in the past) are very rare, it is more than plausible that the relative frequency of collisions shouldn't become too special – in the sense of deviating significantly from the expectation value (2) – unless the initial configuration itself was very special. And yet, this is extremely difficult to *prove*; so difficult, in fact, that, as of to date, the best mathematical results available are valid only for very short times and a very restricted class of particle-interactions.[7]

Moreover, it is important to understand that, unless one considers the thermodynamic limit of an infinitely large system, equation (2) will hold at best *approximately* for *all but a small* set of "bad" initial conditions, that this approximation will get worse with time, and that the approximation is only good enough until it isn't. Eventually, a typical system will exhibit sizable fluctuations out of equilibrium at which point its evolution is no longer adequately described by the Boltzmann equation.

3.3 The *Stoßzahlansatz* as a typicality statement

With all that said, we can now emphasize the fact that the Boltzmann equation and the H-theorem are not an alternative way to explain convergence to equilibrium and the irreversibility of thermodynamic behavior, but rather a concrete exemplification of the explanatory scheme that we have presented before, in more general terms, as the typicality account. Although the micro/macro distinction does not appear as prominently in the formulation of the H-theorem, an essential part of it is that the empirical distribution $f(q, v)$ pertains to a *coarse-grained* description of the system, hence distinguishing a macro-region in phase-space consisting of all microscopic configurations coarse-graining to the same μ-space density. Convergence to equilibrium is then established for *typical initial conditions* with respect to that initial non-equilibrium macro-region. And the equilibrium state – characterized by the Maxwell-distribution to which non-equilibrium distributions typically converge by virtue of the H-theorem – is, as always, distinguished by the fact that it's the one realized by an *overwhelming majority* of all microscopic configurations. As Boltzmann himself beautifully explained:

> *The ensuing, most likely state [...] which we call that of the Maxwellian velocity-distribution, since it was Maxwell who first*

[7] See (Lanford, 1975) and (King, 1975) for the landmark results and (Gallagher et al., 2012) and (Pulvirenti et al., 2013) for recent extensions to more general potentials.

> *found the mathematical expression in a special case, is not an outstanding singular state, opposite to which there are infinitely many more non-Maxwellian velocity-distributions, but it is, to the contrary, distinguished by the fact that by far the largest number of all possible states have the characteristic properties of the Maxwellian distribution, and that compared to this number the amount of possible velocity-distributions that deviate significantly from Maxwell's is vanishingly small.* (Boltzmann, 1896a, p. 252, translation by the authors)

Despite the common focus on the *Stoßzahlansatz*, there is a compelling case to make that the tendency to equilibrium is by all means *explained* by the dominance of the equilibrium state. (Although it will not appear among the premises of the H-theorem, nor necessarily as an explicit part of the proof!) The explanatory role of the Stoßzahlansatz is then somewhat subsidiary to this insight, namely to express the fact that it's thus the "most likely" evolutions carrying a non-equilibrium distribution into equilibrium.

Finally, we understand that the *irreversibility* of the Boltzmann equation (as an effective description of a system's macro-evolution) is – as it cannot be otherwise – a consequence of the fact that non-equilibrium configurations converging to equilibrium are *typical* with respect to the corresponding "macrostate", whereas microscopic configurations leading to the time-reversed evolution are *atypical* with respect to all equilibrium configurations, i.e. all microstates coarse-graining to $f_{eq}(q, v)$.

One will often encounter the claim that the irreversibility of the Boltzmann equation is a result of the *Stoßzahlansatz* being an explicitly time-asymmetric assumption (e.g. Uffink, 2008; Price, 1996, 2002). This is not correct. Of course, it is hard to see how a time-asymmetric assumption about collisions described by reversible microscopic laws could be justified, but Boltzmann's arguments contain no such questionable ploys. The assumption of molecular chaos breaks the time-symmetry only in the obvious (and necessary) sense that it applies to the thermodynamic evolution but not to the reversed motion; but this does not mean that any time-asymmetry is smuggled into the derivation of the H-theorem in addition to the one introduced by the assumption of a non-equilibrium initial distribution.

This misunderstanding, we believe, is mostly based on the failure to recognize molecular chaos, respectively the *Stoßzahlansatz*, as a *typicality statement*. For *typical* initial conditions, eq. (2) is equally valid for the time-evolution in *both temporal directions*. However, the microscopic configurations that have evolved from a state of lower entropy are *ipso facto* atypical with respect to their evolution in the reversed (past) time direction.

To put it differently, if the assumption of molecular chaos is justified in the sense explained before, it will hold for typical initial configurations realizing a non-equilibrium distribution, for which the H-theorem thus asserts convergence of the distribution function to a Maxwellian distribution (towards the future as well as towards the past) and it will also hold for typical equilibrium configurations, for which the H-theorem thus asserts that the equilibrium distribution is stationary. There is no reason, however, why it must hold for those equilibrium configurations that are the time-reversal of states that have just evolved from non-equilibrium, which are, after all, a vanishingly small subset of the equilibrium region. And we know, of course, that it doesn't, that those states are precisely contained in the set of bad configurations for which the particles are correlated in such a way as to undergo a macro-evolution of decreasing entropy (increasing H) that cannot be described by the Boltzmann equation. And we also know that the atypicality of these states (with respect to their evolution in one temporal direction) is explained by, or at least a necessary consequence of, the fact that the system is assumed or constrained or observed to be in a special (i.e. non-equilibrium) state at one particular moment in time.

The only deeper question that may be left is why the Boltzmann equation is in fact *relevant*, i.e. why it is a good description of an *actual* gas in our *actual* world. To understand the answer to this question is thus to appreciate the meaning and relevance of typicality statements.

4 Typicality and the status of macroscopic laws

4.1 The 'logic' of typicality statements

One of the hurdles that may have stood in the way of appreciating Boltzmann's contribution and the relevance of typicality is the fact that Nagelian schemes of reduction and the related *deductive-nomological models* of physical explanation did not quite capture the subtleties of Boltzmann's arguments.[8] According to these often criticized yet very persistent theories, a microscopic explanation of the second law of thermodynamics – respectively

[8]See (Dizadji-Bahmani et al., 2010) for a recent defense of Nagelian reduction. On typicality, see, e.g., (Maudlin, 2007; Bricmont, 1995; Dürr, 2009; Goldstein, 2012; Zanghi, 2005).

a reduction by the microscopic theory – must be a *derivation* of the macroscopic law from the microscopic laws plus suitably specified auxiliary assumptions or "circumstances" in which the macroscopic law is supposed to hold. There is a certain sense in which, in the end, we will concur with this characterization. But first, we want to emphasize one of the more problematic aspects of this view, which is that an understanding of the relationship between the macroscopic regularity and the underlying microscopic laws in purely logical terms misses the crucial role that *initial conditions* play in the explanation of a macroscopic phenomenon.

For what is it to *derive* the thermodynamic behavior of, let's say, a gas from the microscopic laws of motion? Is it to show that there exists at least one microscopic configuration for which the gas will relax to equilibrium? Is it to show that it will happen for *all* possible (non-equilibrium) configurations? The insufficiency of the first statement and the falsehood of the second must severely question the adequacy of purely deductive schemes of explanation. For suppose we wanted to account for the thermodynamic behavior of a certain type of physical system by a scheme of the form $\forall x\big(F(x) \Rightarrow G(x)\big)$, where x ranges of all possible realizations of the corresponding microscopic model and the predicate G is a suitable formulation of "showing effectively/approximately thermodynamic behavior". Then the antecedent $F(x)$ would have to contain a clause more or less equivalent to the statement "The initial conditions of the system x are such that $G(x)$". But then the deduction becomes too trivial to be relevant. *Of course* there exist initial conditions for which the gas will expand. There are also initial conditions for which the gas will contract. And initial conditions for which the gas will transform into a banana. In other words, for a system x with sufficiently many degrees of freedom and sufficiently non-trivial dynamics it is practically *always* possible to maintain that it has the (macroscopic) property G because the initial conditions were such that $G(x)$. The only thing that can provide explanatory value in this context is the assertion of *typicality*, i.e. the assertion that G is not a feature of certain *special* initial conditions, but a physical fact that would arise from *almost any* initial condition. This is also to assure that the explanatory work is done, as much as possible, by the fundamental *laws*, rather than by some fine-tuned arrangement of microscopic degrees of freedom.[9]

Note however that the relevant statement is now, logically and syntactically, a proposition about G rather than a proposition about any particular

[9]Thanks to Jenann Ismael for this insight.

x. The "logic" of the statistical explanation of the second law is thus not to state a set of (statistical) assumptions about an *individual* system from which to infer its thermodynamic behavior, but to spell out a physical account that *grounds* the explanation of thermodynamic behavior in the notion of typicality.

4.2 Typicality vs. probability

Indisputably, the common way of speaking is not to assert that a macroscopic feature G is typical, but to say $G(x)$ is *very likely* or that we infer $G(x)$ *with high probability.* Such a probabilistic statement must, however, raise two additional questions: a) what is it supposed to mean? and b) how did we accomplish the feat to derive a probabilistic result from deterministic microscopic laws? We cannot discuss here in detail how the different "interpretations" of the concept of probability (subjectivist, frequentist, etc.) fare in the context of our discussion, but we want to shed light on a few general points to capture the intricacy of the issue.

First, it would seem rather odd (and detached from scientific practice) if, in order to account for the second law of thermodynamics, we would have to add to the mechanical laws a quantitative assumption about the distribution of initial conditions of boxes of gas, or the like, that we find in our universe.

Second, the fact that we are generally ignorant about the exact microstate of a system is true, but largely irrelevant. It is absurd to think that the validity of the second law of thermodynamics could in any way depend on what we know or believe or are able to observe.

Finally, if we are serious about our commitment to argue within the paradigm of a particular deterministic theory, we have to take it to the conclusion that there is nothing more random about the physical processes that give rise to subsystems in non-equilibrium states than about the entropy-increasing processes going on within these subsystems, once they are suitably isolated from their environment. Eventually one has to wonder why it is true *as a matter of fact* that whenever someone prepares a gas in a low-entropy state, it never ends up in one of the "bad" microscopic configurations for which the gas would contract rather than expand. And then one has to take seriously the fact that an act of "preparation" is itself a physical process, following the same set of physical laws, with its outcome determined by suitably specified initial conditions. Why are *these* initial conditions always good ones, then? To defer the source of randomness to the outside, from the box of gas to the shaky hands of the experimentalist or to exterior perturbations preventing the subsystem from being perfectly isolated, is just to pass the buck. But the buck must stop, eventually, with the universe itself. For the universe is what it is, it exists once and only once, there is nothing before and nothing outside. And we either live in a universe that obeys the second law of thermodynamics (on cosmological scales and, with the possibility of very rare exceptions, in its branching sub-systems) or we don't.

All that said, what is the difference between a statement of probability and a typicality statement, and why is typicality the more appropriate concept in this context?

For one thing, contrary to the conventional use of probabilities, typicality is not a *quantitative* concept. The role of the typicality measure is only to realize and give precise meaning to the notion of "almost all" or "the overwhelming majority of" initial conditions and although it is common and convenient and natural to use the Liouville measure, at least in the context of classical mechanics, many different measures would yield the same notion of typicality.[10] In particular, we are not committed to giving meaning to the exact number that the typicality measure assigns to every (measurable) subset of phase-space. The only "probabilities" that are meaningful in this context are 1 (or those close to 1) and 0 (or those close to 0) indicating what Bernoulli called *moral certainty* and *moral impossibility*.[11]

Furthermore, in making a typicality statement, we do not commit ourselves to talking about actual or hypothetical *ensembles* of systems, nor do we use probabilistic concepts to express our "guess" – in terms of *information* or *knowledge* or *believe* – about a system's actual microstate. A typicality statement refers to nothing more and nothing less than the fact that a certain (coarse-grained/macroscopic) property or behavior of a physical system is typical according to the microscopic laws, i.e. that it's the kind of property or behavior that our fundamental theory predicts for an overwhelming

[10]On the other hand, many measures would yield a different notion of typicality. One can think, for instance, of singular measures, concentrated on a single point in phase-space. Such a measure may even turn out to be stationary, in case that this particular microstate happens to be a stationary point of the dynamics. So why not take such a measure to define "typicality", meaning that a property is typical if and only if it is instantiated by this one particular configuration? We trust the reader to answer this question for himself.

[11]See (Bernoulli, 1713). Such typicality statements can be understood in the sense of *Cournot's principle*, which is one of the basic principles underlying the philosophy of Kolmogorov's *Grundbegriffe*, but also stands in the philosophical tradition of great mathematicians such as Emile Borel, Maurice Fréchet or Paul Lévy. See (Shafer and Volk, 2006) for a beautiful essay on this topic.

majority of microscopic configurations compatible with appropriately specified (macroscopic) boundary conditions: Typically, a coin tossed repeatedly for a large number of times will land about as often on *heads* as on *tails*. Typically, an ice cube at room temperate will melt. According to the laws of quantum mechanics, a collection of point-particles shot successively through a double-slit will typically (though not necessarily) display an interference pattern when registered on a screen behind the slits. According to classical mechanics, it typically won't (although it possibly might).

A typicality statement is thus an objective physical fact, *in principle* derivable from the fundamental (microscopic) laws that we take as the basis of our considerations. (It is a fact that, by the way, even Laplace's demon should care about, to the degree that he cares about physics.) But what exactly is it a fact about? Well, typicality is, first and foremost, the answer to the question that stood at the very beginning of our discussion, namely: what is the connection between the macroscopic regularities that physics is supposed to account for and the underlying microscopic laws. Another way to put it is to ask: What is the nomological status of the "macroscopic laws"?

4.3 Typicality and the status of macroscopic laws

Philosophically, the truly remarkable yet often unacknowledged aspect about the probabilistic character of thermodynamic laws is not the way in which laws that once have been thought to be exact turn out to be merely "approximately true", but the way in which the regularities expressed by these laws turn out to be *contingent* rather than *necessary* truths. In other words, if we accept the microscopic laws as fundamental, we have to accept that the so-called "macroscopic laws", even in an approximate or statistical sense, are in fact *no laws at all* in that they lack the status of nomological necessity. For all we know, the initial conditions of our universe (conceived as a Newtonian universe) could have been such that systems, prepared or created in a low-entropy state, would regularly end up on one of the "bad" trajectories that undergo an anti-thermodynamic evolution of decreasing entropy. That is to say that there are possible Newtonian universes in which gases are regularly found to contract rather than expand, in which heat does sometimes flow from a colder to a hotter body and in which macroscopic objects such as balls and chairs and tables occasionally jump up in the air (while cooling off accordingly to account for the conservation of energy) simply because a large number of particles happen to move in the same direction at the same time.

In these *counterfactual* but *nomologically possible* universes, it is simply *not true* that such events are very unlikely, because they happen "all the time".[12]

And yet, we would insist, it's more than a mere contingency, more than a *factum brutum* that our universe is not like that. And indeed, our physical theory has more to say here – fortunately without assigning us the impossible task of determining the *actual* boundary conditions of our universe – for it tells us that the initial conditions of a Newtonian universe would have to be *exceedingly special* to give rise to subsystems violating thermodynamic laws as more than astronomically rare exceptions. Thermodynamic laws, in other words, are *statistical regularities of typical universes.* And it is this characterization, we suggest, that specifies their connection to the underlying microscopic laws and grounds their own law-like status.

Kripke (1980) famously explained the difference between logical and nomological supervenience by the following metaphor: B-properties supervene logically on A-properties if, after fixing the A-properties of the world, there was nothing else God could (or needed to) do for fixing the B-properties. The A-properties, we say, *logically entail* the B-properties. In case of a nomological supervenience, however, God, after making sure of the A-facts, still had some work to do for making sure of the B-facts by determining *laws of nature* relating B-properties to A-properties. Going one step further, we can say: a property of our world that is *typical* for these laws, is a fact or regularity for which God, after fixing the laws of nature and the fundamental ontology of the world, still had a *little bit* of work to do in choosing appropriate initial conditions for the universe. However, while almost any possible choice (compatible with the relevant macroscopic constraints) would have been fine to make sure that this property is instantiated, God would have had to be utterly meticulous – and maybe somewhat malicious - to arrange the initial configuration of the universe in such that it isn't.

Turning back to the 'second law', we have to note one subtlety in connection with the Past Hypothesis (see Section 2). According to the Past Hypothesis, the initial *macrostate* of our universe was a very *special* one, marking the low-entropy end of the thermodynamic arrow of time. However, with respect to this macrostate, the initial *microstate* of the universe was *typical* (in regard to its future evolution), thus explaining the increase

[12] Of course, among all possible Newtonian universes there will be many with no thermodynamic arrow and no interesting structures at all, but here, to make a point, we consider universes that are hospitable to intelligent life, while the second law of thermodynamics fails to hold in branching systems just as often as to make a fool out of physicists.

of entropy in the universe as a whole and in any of its branching subsystems. All in all, there is no contradiction, but a clear *tension* between the typicality account and the Past Hypothesis. The resolution of this tension is considered by many as one of the most profound problems of modern physics.[13]

What else is left to say? Not much, we believe. To understand that a certain regularity is typical and yet to wonder why it is that we observe this regularity in nature (and why we should expect this regularity to persist in the future), is to ask why our universe is typical, i.e. why it is, in this particular respect, like the overwhelming majority of all possible universes instantiating the same set of fundamental physical laws. And while we don't know how to answer, except maybe with Einstein's bon mot that "God is subtle, but he is not malicious", the very question seems to us utterly uncompelling. Explanations have to end somewhere. If we can establish that a certain property is typical for a particular kind of system, this should elevate any sense of mystery or puzzlement as to why we find such systems instantiating the respective property. Hence, we should consider the phenomenon to be reasonably and conclusively *explained* on the basis of the microscopic theory. Similarly, if we can establish that a macroscopic feature or regularity is typical for a certain kind of system, we should by all reasonable means expect to find this feature realized in a given system of the said kind. Hence, it constitutes a *prediction* of the microscopic theory.

In this fashion, typicality statements figure in a *way of reasoning* about nature. In fact, since the situation in which we find ourselves towards the world is necessarily one in which all we can ever hope to know about the world's state is compatible with a plurality of fundamental (microscopic) matters of fact, the relevant *explanatory* and *behavior guiding* statements that we can extract from the fundamental laws of physics are virtually always results about *typical solutions* of their equations of motion.

Finally, we shall emphasize again that a typicality reasoning is a non-deductive reasoning. *Logically*, the fact that something has been shown to be typical doesn't imply anything about any *particular* instance. In other words, it is always *possible* for a particular system – and ultimately our universe – to be atypical in the relevant respect. But facts that strike us as atypical are

[13]See, for instance, (Penrose, 1999) and his "Weyl curvature hypothesis" as a proposal for an additional law restricting the initial state of the universe, but also (Callender, 2004) arguing from a Humean perspective *against* the need for further explanation of the Past Hypothesis. See (Carroll, 2010) for a very nice discussion of the problem as well as (Carroll and Chen, 2004) for an attempt to dispose of the Past Hypothesis altogether.

usually the kind of facts that cry out for *further* explanation. This is why a Casino manager has not just economic interest but reasonable grounds to suspect cheating if a player hits three jackpots in a single night. And this is why scientific practice would eventually require us to revise our theory and look for different laws, rather than endorsing an explanation of empirical data based on special initial conditions or, if you will, a streak of bad luck. In the end, it is not logically but epistemically inconsistent to accept a certain physical theory and accept at the same time that our universe is somehow an atypical model of that theory, for this would undermine any reasons to endorse the theory in the first place.[14]

5 Reply to critics

5.1 Missing the point of typicality

Despite the many subtleties involved in the concept of typically, we believe this way of reasoning to be very natural and intuitive and very much in line with common scientific practice. Nevertheless, it seems to us that quite a lot of misunderstandings concerning Boltzmann's statistical mechanics are actually failures to appreciate the "logic" of typicality statements. One of the most common mistakes, in fact, is simply to miss the difference between a typicality statement and a statement about particular instances. Consider, for instance, the objection of Roman Frigg in reply to Goldstein (2001):

> *Goldstein suggests that a system approaches equilibrium simply because the overwhelming majority of states in Γ_E are equilibrium microstates [...]. This is wrong. If a system is in an atypical microstate [...], it does not evolve into a equilibrium microstate just because the latter are typical; typical states do not automatically function as attractors. (Uffink, 2007, 979–980) provides the following example. Consider a trajectory $x(t)$, i.e. the set $\{x(t) = \phi_t(x(t_0)) \mid t \in [t_0, \infty)\}$, a set of measure zero in Γ_E. Its complement, the set $\Gamma_E \setminus x(t)$ of points not laying on $x(t)$, has measure one. Hence the points on $x(t)$ are atypical while the ones not on $x(t)$ are typical (with respect to Γ_E, μ, and the property 'being on $x(t)$'). But from this we cannot conclude that a point on $x(t)$ eventually has to move away from $x(t)$ and end up in $\Gamma \setminus x(t)$; in fact the uniqueness theorem for solutions tells*

[14] As was put so nicely by Mathias Frisch (private communication).

> *us that it does not. The moral is that non-equilibrium states do not evolve into equilibrium states simply because there are overwhelmingly more of the latter than of the former, i.e. because the former are atypical and the latter are typical.* (Frigg, 2009, pp. 8–9).

Of course, no one is claiming, in the naive sense implied by Frigg, that any *specific trajectory* will move to equilibrium "simply because" equilibrium states are typical – just as no one claims that any specific lottery ticket must lose "simply because" losing lottery tickets are typical. In the alluded sense, a lottery ticket loses simply because someone picked the wrong numbers and a system converges to equilibrium simply because its micro-evolution carries the microscopic configuration into an equilibrium state. The relevant assertion here is that the regions of phase-space that do not correspond to the thermodynamic equilibrium are extremely special. And the claim is then that solution-trajectories that wander around in phase-space, yet remain confined, for an extensive amount of time, to those extremely special regions of phase-space, will turn out to be themselves extremely special. And this is to say, in other words, that *typical* initial conditions in non-equilibrium will evolve into equilibrium and that typical equilibrium states will remain in (or close to) equilibrium over very long periods of time.

So what is the point of the "counterexample" formulated by Jos Uffink that made such an impression on Frigg? It's obviously correct that a solution $x(t)$ of the equations of motion will never enter the phase-space region $\Gamma_E \setminus x(t)$ despite the fact that it has measure 1. *Typical* solutions, however, will. In fact, it follows from the "uniqueness theorem" that *every* other solution (with the same total energy) lies *entirely* in the set $\Gamma_E \setminus x(t)$. So, leaving aside the fact that this artificially crafted region of phase-space is of no physical interest whatsoever, it is not clear what this example is actually supposed to demonstrate. With all due respect, the debate seems a bit like people trying to explain that a typical lottery ticket will fail to win the jackpot because of the huge number of combinations that could be drawn, and Frigg and Uffink running around with a winning lottery ticket in order to disprove them.

5.2 A comment on the 'measure zero problem'

If Uffink's example works at all, then as another instance of the so-called "measure zero problem", which is basically the observation that, as soon as one goes to a more *fine-grade* description, any physical system is found to be atypical with respect to some (more or less natural) properties. In particular, for a continuous state-space and a nonsingular measure, the *actual* microscopic configuration and, as we just noted, even the entire trajectory of a system will constitute a set of measure 0. Although this observation receives ongoing interest and is often presented as a serious challenge to typicality arguments (cf. Frigg, 2009, p. 23; Sklar, 1993), we don't think that it causes much of an embarrassment for the reasoning we presented.[15]

There are facts and regularities that can be explained on the basis of the microscopic laws by virtue of being typical (like the frequency of 'head' and 'tail' in long series of coin-tosses being approximately 50 : 50). There are contingent facts about physical systems that are not typical, but can be explained in a different sense – usually by tracing them back to other (even more) special states of affair. For instance, the state of our office is certainly atypical with respect to the exact distribution of objects on the desk, but we can tell some sort of causal story about how a used coffee mug ended up near the keyboard and how the battered blue book came to lie on top of the heavier red one. And finally there are facts like the one that a trajectory through some physical state-space will never cross its complement – which do not require further explanation, but are well-suited for creating confusion where none is due.

5.3 The role of the dynamics

A different objection to the typicality account that can be found in the philosophical literature is that it fails to make precise the dynamical assumptions on which the argument rests (Uffink, 2008; Frigg, 2009, 2011; Frigg and Werndl, 2012). Frigg and Werndl (2012) even go as far as declaring that the typicality account is "mysterious" because the "connection with the dynamics" is unclear (p. 918). Jos Uffink (2008) writes on a similar note (as a conclusion to his "counterexample" recited by Frigg and discussed above):

> *[I]n order to obtain any satisfactory argument why the system should tend to evolve from non-equilibrium states to the equilibrium state, we should make some assumptions about its dynamics. In any case, judgments like 'reasonable' or 'ridiculous' remain partly a matter of taste. The reversibility objection is a request for mathematical proof (which, as the saying goes, is something*

[15] Thanks to Tim Maudlin for very helpful discussions on this issue.

> *that even convinces an unreasonable person).* (Uffink, 2007, p. 61)

We have both very much and very little to say to this objection. First and foremost, we can emphasize that the typicality account is an explanation or explanatory scheme – not a proof. To reject it for a lack of mathematical rigor is thus to miss the point entirely. It lies in the nature of the problem that rigorous mathematical results about systems with roughly 10^{23} degrees of freedom are very hard to come by. And even in suitable thermodynamic limits (where the number N of particles goes to infinity and other quantitates in the microscopic model scale accordingly), proving convergence to equilibrium for a more or less realistic model remains an extremely difficult and largely unresolved problem of mathematical physics. That said, while everybody is entitled to his epistemic standards, there are usually good reasons to settle for physical explanations that are conclusive enough to convince a reasonable person. If, as an explanation of a macroscopic phenomenon, we accepted nothing short of rigorous mathematical proof, the atomic hypothesis would yet have to earn its merits. What is less a matter of personal standards, however, is the reference to the reversibility objection. For the reversibility objection, we must insist, is not so much a "request for mathematical proof" as a request for a conclusive explanation of macroscopic irreversibility – or so it was in 1876. Boltzmann provided a conclusive explanation soon after and we now have a very good understanding of how irreversible macroscopic behavior can arise from reversible microscopic dynamics. Moreover, even if Uffink is not satisfied with Boltzmann's answer, it's unclear what kind of dynamical assumption could help him, since one thing the dynamics certainly are, by assumption, is reversible.

All in all, it's hard to discern what exactly Uffink is confused about and what precisely he's objecting to. Frigg and Werndl, very much to their credit, state more clearly what they have in mind:

> *In recent years several proposals have been put forward, which aim to justify (something akin to) TD-like [thermodynamic-like] behaviour in terms of typicality [...]. This programme is on the wrong track. [...] Not all phase flows lead to TD-like behaviour (for instance, a system of harmonic oscillators does not). So the phase flows that lead to TD-like behaviour are a non-trivial subclass of all phase flows on a given phase space, and the question is how this class can be characterised. [...] What we need*

> *is a non-trivial specification of a property that only those flows that give raise to TD-like behaviour possess.* (Frigg and Werndl, 2011, pp. 4–5)

These demands, however, are much less reasonable than it might first seem. To start with, it is important to keep in mind that while in statistical mechanics we will sometimes discuss many-particle systems on the level of dynamical systems, what we care about, in the end, is not dynamical system theory. In particular, we do not care about measure-preserving flows *in general*, but always have in mind a phase-flow generated by a huge number of interacting particles constituting a particular physical system of interest.

Obviously, whenever we study a specific model, whether it possesses the appropriate characteristics, i.e., whether it describes a gas rather than a fluid (or nothing interesting at all) and whether it exhibits the right thermodynamic behavior, will in the end depend on the Hamiltonian, comprising the particle interactions and determining the system's time-evolution. Also, one should keep in mind that dynamical considerations, to a certain degree, are already reflected in the partitioning of a system's phase-space and the determination of the volume (respectively entropy) corresponding to the various macro-regions. In particular, the equilibrium state can look very differently depending on the broad characteristics of the microscopic interactions (e.g., for a drop of ink compared to a drop of oil in water).

On the other hand, we understand from Boltzmann's analysis that, once this stage is set, the explanation of thermodynamic behavior is extremely robust against the details of the microscopic model, precisely because it doesn't hinge on any narrowly-conceived properties of the dynamical system or the interaction potentials. In particular, the *explanatory work* is almost entirely done by the dominance of the equilibrium state and the notion of typicality, without the need to emphasize special features of the dynamics. The reason is simply that once we understand that the non-equilibrium region of phase-space is vanishingly small compared to the equilibrium region, we see that there is nothing special or remarkable about the dynamics for which a typical set of solutions, starting in a non-equilibrium region, will quickly spread over the equilibrium region and for which equilibrium configurations will stay in equilibrium for most of the time. If you throw a rubber duck into the Atlantic Ocean, what do you need to know about oceanic currents in order to explain and predict and understand that it will almost certainly spend most (if not all) of the time outside the region where the Titanic sank? It is precisely the generality of Boltzmann's argument that makes it so powerful, giving us an understanding of thermodynamic behavior as a virtually universal feature of macroscopic systems.

All that said, we share neither Frigg's and Werndl's interest in characterizing general phase-flows with respect to their thermodynamic behavior, nor their expectation that it should be possible to state necessary and sufficient criteria for convergence to equilibrium in terms of precise mathematical properties characterizing the dynamics of roughly 10^{23} interacting particles. In particular, we fail to see anything of physical or philosophical interest in examples such as that of a system of uncoupled harmonic oscillators not exhibiting thermodynamic behavior.[16] In fact, the latter claim is not entirely accurate. If one considers a system of oscillators with various but similar frequencies, there is an interesting sense in which it can converge to equilibrium, namely from a state in which the oscillators are more or less in phase into a state in which they aren't.[17] Presumably, what the authors mean to say is that if we consider uncoupled harmonic oscillators as a model for the gas, the argument for its convergence to equilibrium won't go through. And presumably, the authors are not actually preoccupied with the question why it is that a collection of wiggling particles will not spread over a given volume, but mean to demonstrate that the typicality account must be incomplete or inconclusive because its conclusion does not follow from its premises. But to argue like this is to misunderstand the nature of the explanation in the first place, which has never been about stating a set of mathematical assumptions from which to prove thermodynamic behavior in the abstract.

Indisputably, concerning the microscopic derivation of the second law, very little is on firm mathematical ground. This is just a fact about the current status of science. It is a fact that one might be unhappy about and it is certainly a fact that will continue to motivate further research. However, it is of utmost importance to understand that, contrary to what some commentators have suggested, the difference between the explanatory scheme that we have presented and a more rigorous *proof* of the second law is not some secret ingredient like a dynamical assumption that proponents of the typicality account have missed to specify, but a heap of very hard, very technical work in mathematical physics. Good physics and good philosophy of physics, on the other hand, is also about appreciating where our understanding of an issue depends on rigorous formalization and technical proof and where it doesn't.

[16] Which, by the way, is a common line of argument.

[17] See also the analysis of a system of anharmonic oscillators in (Bricmont, 2001).

References

[Albert, 2000] Albert, D. (2000). Time and Chance. (Cambridge: Harvard University Press)

[Bernoulli, 1713] Bernoulli, J. (1713). Ars conjectandi, opus posthumum. Accedit Tractatus de seriebus infinitis, et epistola gallicé scripta de ludo pilae reticularis. (Basel: Thurneysen Brothers). Reprinted: Bernoulli, J. (2006). The art of conjecturing. (Baltimore: The John Hopkins University Press)

[Boltzmann, 1896a] Boltzmann, L. (1896). Vorlesungen über Gastheorie. (Leipzig: Verlag v. J. A. Barth, Leipzig). Nabu Public Domain Reprints.

[Boltzmann, 1896b] Boltzmann, L. (1896). Entgegnung auf die wärmetheoretischen Betrachtungen des Hrn. E. Zermelo. Wiedemann's Annalen, 57, 773-784.

[Bricmont, 1995] Bricmont, J. (1995). Science of Chaos or Chaos in Science? Physicalia Magazine, 17, 159-208.

[Bricmont, 2001] Bricmont, J. (2001). Bayes, Boltzmann and Bohm: Probabilities in Physics. (In J. Bricmont, D. Dürr, et al. (Eds.), Chance in Physics. Foundations and Perspectives (pp. 3-21). Berlin: Springer.)

[Callender, 2004] Callender, C. (2004). There is No Puzzle about the Low Entropy Past. (In C. Hitchcock (Ed.). Contemporary Debates in Philosophy of Science (pp. 240-255). London: Blackwell.)

[Carroll and Chen, 2004] Carroll, S., & Chen, J. (2004). Spontaneous Inflation and the Origin of the Arrow of Time. ArXiv: hep-th/0410270.

[Carroll, 2010] Carroll, S. (2010). From Eternity to Here. The Quest for the Ultimate Theory of Time. (USA: Dutton, Penguin Group)

[Davies, 1977] Davies, P. C. W. (1977). The Physics of Time Asymmetry. (Berkeley: University of California Press)

[Dizadji-Bahmani et al., 2010] Dizadji-Bahmani, F., Frigg, R. & Hartmann, S. (2010). Who's Afraid of Nagelian Reduction? Erkenntnis, 73, 393-412.

[Dürr, 2009] Dürr, D. (2009). Bohmian Mechanics. (Berlin: Springer)

[Ehrenfest, 1911] Ehrenfest, P. und T. (1911). Begriffliche Grundlagen der statistischen Auffassung in der Mechanik. Enzyklopädie der mathematischen Wissenschaften mit Einschluss ihrer Anwendungen, Band 4, 3-90.

[Feynman, 1967] Feynman, R. (1967). The Character of Physical Law. (Cambridge: The M.I.T. Press)

[Frigg, 2009] Frigg, R. (2009). Typicality and the Approach to Equilibrium in Boltzmannian Statistical Mechanics. Philosophy of Science, 76, 997-1008.

[Frigg, 2011] Frigg, R. (2011). Why Typicality Does Not Explain the Approach to Equilibrium. (In M. Suárez (Ed.), Probabilities, Causes and Propensities in Physics (pp. 77-93). Dordrecht: Springer.)

[Frigg and Werndl, 2011] Frigg, R., & Werndl, C. (2011). Explaining Thermodynamic-Like Behaviour In Terms of Epsilon-Ergodicity. Philosophy of Science, 78, 628-652 .

[Frigg and Werndl, 2012] Frigg, R., & Werndl, C. (2012). Demystifying Typicality. Philosophy of Science, 79, 917-929.

[Gallagher et al., 2012] Gallagher, I., Saint Raymond, L., & Texier, B. (2012). From Newton to Boltzmann: the case of short-range potentials. Preprint: ArXiv: 1208.5753v1 [math.AP].

[Goldstein, 2001] Goldstein, S. (2001). Boltzmann's Approach to Statistical Mechanics. (In J. Bricmont, D. Dürr, et al., Chance in Physics. Foundations and Perspectives (pp. 39-54). Berlin: Springer.)

[Goldstein, 2012] Goldstein, S. (2012). Typicality and Notions of Probability in Physics. (In Y. Ben-Menahem, & M. Hemmo (Eds.), Probability in Physics. The Frontiers Collection (pp. 59-71). Berlin: Springer.)

[Goldstein and Lebowitz, 2004] Goldstein, S. & Lebowitz, J. (2004). On the (Boltzmann) entropy of non-equilibrium systems. Physica D: Nonlinear Phenomena, 193, Issues 1-4, 53-66.

[King, 1975] King, F. (1975). BBGKY hierarchy for positive potentials. Dissertation, University of California at Berkeley.

[Kripke, 1980] Kripke, S. (1880). Naming and Necessity. (Oxford: Blackwell)

[Lanford, 1975] Lanford, O. E. (1975). Time Evolution of Large Classical Systems. (In J. Moser (Ed.), Lecture Notes in Physics, Vol. 38 (pp. 1-111), Berlin: Springer.)

[Lebowitz, 1981] Lebowitz, J. (1981). Microscopic dynamics and macroscopic laws. Annals New York Academy of Sciences, 220-233.

[Lebowitz, 1993] Lebowitz, J. (1993). Macroscopic laws, microscopic dynamics, time's arrow and Boltzmann's entropy. Physica A, 194, 1-27.

[Maudlin, 2007] Maudlin, T. (2007). What could be objective about probabilities? Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics, 38, Issue 2, 275-291.

[Penrose, 1999] Penrose, R. (1999). The Emperor's New Mind. (Oxford: Oxford University Press)

[Price, 1996] Price, H. (1996). Time's Arrow & Archimedes' Point. New Directions for the Physics of Time. (New York: Oxford University Press)

[Price, 2002] Price, H. (2002). Burbury's Last Case: The Mystery of the Entropic Arrow. (In C. Callender (Ed.), Time, Reality & Experience (pp. 19-56). Cambridge: Cambridge University Press.)

[Pulvirenti et al., 2013] Pulvirenti, M., Saffiro, C., & Simonella, S. (2013). On the validity of the Boltzmann equation for short range potentials. Preprint: ArXiv: 1301.2514v1 [math-ph].

[Schwartz, 1992] Schwartz, J.: (1992) The pernicious influence of mathematics on science. (In M. Kac, G.-C. Rota, J. and Schwartz (Eds.), Discrete Thoughts (pp. 19-25). Boston: Birkhäuser.)

[Shafer and Volk, 2006] Shafer, G., & Volk, V. (2006). The Sources of Kolmogorov's Grundbegriffe. Statistical Science, 21, No. 1, 70-98.

[Sklar, 1973] Sklar, L. (1973). Statistical Explanation and Ergodic Theory. Philosophy of Science, 40, No. 2, 194-212.

[Spohn, 1991] Spohn, H. (1991). Large Scale Dynamics of Interacting Particles. (Berlin: Springer)

[Uffink, 2007] Uffink, J. (2007). Compendium of the foundations of classical statistical physics. (In J. Butterfield, J., & J. Earman (Eds.), Handbook for the Philosophy of Physics (pp. 923-1047). Amsterdam: Elsevier.)

[Uffink, 2008] Uffink, J. (2008). Boltzmann's Work in Statistical Physics. The Stanford Encyclopedia of Philosophy.

[Villani, 2002] Villani, C. (2002). A review of mathematical topics in collisional kinetic theory. (In S. Friedlander, & D. Serre (Eds.), Handbook of mathematical fluid dynamics, Vol. 1 (pp. 71-305). Amsterdam: Elsevier.)

[Zanghì, 2005] Zanghì, N. (2005) I fondamenti concettuali dell'approccio statistico in fisica. (In V. Allori, M. Dorato, F. Laudisa, & N. Zanghi (Eds.), La Natura Delle Cose. Introduzione ai Fundamenti e alla Filosofia della Fisica (pp. 139-228). Roma: Carocci.