

Typicality, Irreversibility and the Status of Macroscopic Laws

Dustin Lazarovici*, Paula Reichert†

*Mathematisches Institut, Ludwig-Maximilians-Universität München
Theresienstraße 39, D-80333 München, Germany.*

Abstract

We discuss Boltzmann’s explanation of the irreversible thermodynamic evolution of macroscopic systems on the basis of time-symmetric microscopic laws, providing a comprehensive presentation of what we call the *typicality account*. We then discuss the connection between this general scheme and the H-theorem, demonstrating the conceptual continuity between them. In our analysis, a special focus lies on the crucial role of typicality. Putting things in wider perspective, we go on to analyze the philosophical dimensions of this concept, explaining the connection between typicality and probability, and demonstrate its relevance for scientific reasoning, in particular for understanding the supervenience of macroscopic laws on microscopic laws. The second part of the paper responds to recent objections against the typicality account that have been raised in the philosophical literature. In particular, the concept of ergodicity, or a variant thereof, named “epsilon-ergodicity”, which has been promoted by some authors as a crucial additional assumption on the dynamics, is shown to be of no use for its intended purpose.

*lazarovici@math.lmu.de

†reichert@math.lmu.de

Contents

1	Introduction	3
2	Boltzmann's statistical mechanics	4
3	Typicality and irreversibility	7
3.1	The combinatorial argument	7
3.2	The measure of typicality	9
3.3	Irreversibility	11
3.4	Past Hypothesis and the thermodynamic arrow	13
3.5	Recurrence	14
4	Proving the second law? Typicality and the H-theorem	16
5	Typicality and the status of macroscopic laws	25
5.1	The 'logic' of typicality statements	25
5.2	Typicality vs. probability	27
5.3	Typicality and the status of macroscopic laws	30
5.4	A comment on the 'measure zero problem'	33
6	The role of the dynamics	36
6.1	The role of mathematics	36
6.2	The role of the dynamics. A reply to critics	38
7	Against (epsilon-)ergodicity	43
8	Conclusion	48

1 Introduction

Over the last two decades, a series of papers by various highly 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 by statistical mechanics (Goldstein 2001, Lebowitz 1993, Bricmont 1995, Penrose 1999). None of these authors took much credit for the presented ideas, each of them rather stressed that he was presenting or recapturing the groundbreaking insights of Ludwig Boltzmann who had shown how to explain and (in some sense) derive macroscopic regularities on the basis of the underlying laws for the microscopic constituents of matter. Nevertheless, reintroducing these ideas to Physicists, Mathematicians and Philosophers proved to be 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.

The present paper originated as a response to various publications by Roman Frigg and Charlotte Werndl (Frigg 2009 and 2011, Frigg and Werndl 2011 and 2012) who established themselves as two of the most outspoken critics of this approach to the foundations of statistical mechanics, though in the end, it grew into a more comprehensive discussion of statistical mechanics, the “second law” and the concept of typicality. In section 2, we will recall the basics of Boltzmann’s statistical mechanics and its explanation of the second law of thermodynamics. Section 3 will elaborate on the details of this explanation, in particular with respect to the role of typicality and the issue of irreversibility, providing a comprehensive presentation of what we call the *typicality account*. In section 4, we will discuss the Boltzmann equation and the H-theorem and clarify their relationship to the typicality account, demonstrating the conceptual unity between these two cornerstones of Boltzmann’s work in statistical mechanics that are often, falsely, viewed as alternatives. Section 5 is the most philosophical part of the paper, discussing the meaning of typicality and its relevance to scientific reasoning, in particular in the context of understanding the precise way in which macroscopic laws can be reduced to microscopic laws. In these first sections, we thus hope to provide a comprehensive and up-to-date account of Boltzmannian statistical mechanics and the central concepts involved. In the last two sections, we will then explicitly address some questions and alleged problems that came up in the philosophical literature (but which, ideally, at that

point of our treatment, will not have remained open in the first place). In section 6, we comment on the role of the dynamics in the typicality account, responding, in particular, to criticisms of Jos Uffink, Roman Frigg and Charlotte Werndl, that we found to be thoroughly misguided. In section 7, we will explain why the concept of *ergodicity*, which is often regarded as crucial to the foundations of statistical mechanics, and, even more so, the concept of *epsilon-ergodicity* promoted by Frigg and Werndl are irrelevant to the explanation of thermodynamic behavior.

2 Boltzmann's statistical mechanics

The discussion to which we want to contribute is concerned with the explanation of the irreversible thermodynamic behavior of macroscopic systems. The term “thermodynamic behavior” 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 (e.g. position and momenta of all the particles constituting the system), its macrostate $M(t)$ is specified in terms of physical variables that characterize the system on macroscopic scales (like the volume it occupies, the pressure of a gas or fluid, the distribution of blue ink in water, 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 generally be realized by a large number of different microstates all of which “look macroscopically the same”. The partitioning of the set of microstates into different macrostates 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 being the position- and p 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 physical systems. However, since this macro-evolution supervenes on the evolution of the microscopic configuration, and since the evolution of the microscopic configuration is determined by precise and unambiguous laws of motion, there is no place for macroscopic laws over and above the microscopic laws, and the empirical regularities expressed in the former must be explained or justified on the basis of the latter. This seems like quite a formidable task, though, as it requires us to reconcile the *irreversibility* of thermodynamic behavior with the *time-reversal symmetry* of the microscopic laws of motion. This symmetry implies that for every solution of the microscopic equations of motion realizing a certain sequence of macrostates, there exists another solution passing through the same set of macrostates in opposite order. And yet, we never observe the spontaneous assembly of a gas in one corner of a box, or the spontaneous unmixing of milk and coffee, or heat flowing from a colder body to a hotter one. However, the formidable task of reconciling this macroscopic irreversibility with reversible microscopic laws was indeed accomplished by Ludwig Boltzmann at the end of the 19th century. Lebowitz, Penrose, Goldstein and Bricmont reminded us how. In doing so, each of them stressed two main insights of the great Austrian physicist:

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 entropy¹ 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 many orders of magnitude, with the region of maximum entropy – by definition the equilibrium region – being by far the largest.² This is also known as Boltzmann's *combinatorial argument*.

The two insights stressed above are the key ingredients in Boltzmann's account of the second law of thermodynamics. What we learn from them is that the thermodynamic behavior that we want to explain is in fact not a feature of certain special micro-evolutions, but rather the kind of macro-evolution that would correspond to almost any generic trajectory through phase-space. Since phase-space consists almost entirely of equilibrium states, the evolution of a system's microstate, starting out in a low-entropy region, would have to be extremely peculiar to *avoid* carrying the system into regions of higher entropy and finally into an equilibrium state, while the evolution of a system starting out in equilibrium would have to be extremely peculiar to carry the microscopic configuration into the vanishingly small non-equilibrium region any time soon. Penrose summarizes the argument as follows:

¹which is necessary to identify S with the thermodynamic Clausius entropy that is an extensive variable of state, i.e. the total entropy of two combined independent systems is supposed to be the sum of the entropies of the individual systems.

²Many authors choose to define equilibrium as a *stationary* macrostate of a system. The problem with this definition is that, in the end, one is talking about the empty set, since it turns out that the macrostates that one used to refer to as equilibrium-states are not exactly stationary, though they *are* states of maximum (Boltzmann-)entropy and as a consequence thereof (as we will see) *for all practical purposes* stationary.

We would seem now to have an explanation for the second law! For we may suppose that our phase-space point does not move about in any particularly contrived way, and if it starts off in a tiny phase-space volume, corresponding to a small entropy, then, as time progresses, it will indeed be overwhelmingly likely to move into successively larger and larger phase-space volumes, corresponding to gradually increasing entropy values. (Penrose 1999, p. 408)

The same reasoning, in a different wording, is also found in the other publications cited above. The authors, as one would expect from scientists of their stature, say no more or less than necessary and if the reader feels like the issue is thus already settled, we would readily agree. Yet we notice that their presentations, aiming for simplicity and generality, spared some of the details and intricacies of the argument and it seems that they have thus left too much room for objections and misconceptions, some understandable, some irritating, all ultimately unnecessary. Lebowitz rightly warned us, quoting Ruelle, that the ideas of Boltzmann are “at the same time simple and rather subtle” (Lebowitz 1993, p. 7). The aim of our paper is to spell out these subtleties, which often remain unspoken, thus addressing the most common objections and clarifying the role of typicality in the argument.

3 Typicality and irreversibility

3.1 The combinatorial argument

Let’s now discuss the details of the “typicality account” sketched above for the concrete example of a gas in a box. This example will accompany us as the paradigmatic case for thermodynamic behavior throughout the entire paper. We thus consider a system of about $N = 10^{23}$ particles, interacting by a repelling, short-range potential – or, in the case of an ideal gas, not interacting at all – 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, that is, we consider a configuration of the particles that looks, macroscopically, like a gas filling out half the volume of the box. What kind of evolution, on the macroscopic scale, should we expect for the gas?

Well, as Boltzmann argued, a simple combinatorial argument shows that *the overwhelming majority* of microstates that the system could possibly

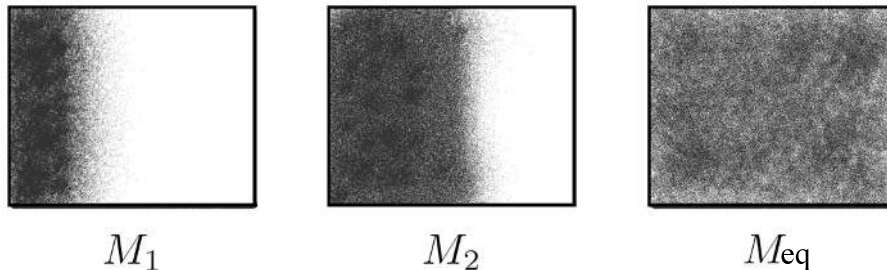


Figure 1: Thermodynamic evolution of a gas

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 phase-space volume corresponding to this *equilibrium macrostate* M_{eq} is about $2^N \approx 10^{10^{23}}$ times larger than the phase-space volume occupied by configurations with substantially lower entropy!³ 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 (though, of course, deterministic) path in the high-dimensional phase-space that should soon end up in the equilibrium region. In other words, we should expect, by all reasonable means, that the evolution of the microscopic configuration will be one that carries the system from M_2 to M_{eq} . However, it is clear (and it was clear to Boltzmann) that there will be microscopic configurations belonging to the macrostate M_2 , configurations, that is, which look like a gas filling the left half of the box, that will *not* show this expected macroevolution, but evolve into a state of *lower* entropy. This is a consequence of the time-reversal invariance of the microscopic laws. For if we consider a macrostate of even lower entropy, M_1 , we have to conclude that for every solution of the microscopic laws of motion corresponding to the macro-evolution from M_1 to M_2 (which is the evolution that we would now predict by an analogous reasoning), there exists a solution carrying an initial microstate in Γ_{M_2} into the macro-region Γ_{M_1} – we just have to take the microstate in Γ_{M_2} that has evolved from Γ_{M_1} and reverse the momentum of every particle! This was the content of Loschmidt's

³Comparing, for instance, the possible equilibrium configurations with all possible configurations that coarse-grain to the macrostate M_2 , we see that every single particle has roughly twice the volume that it could be in.

reversibility objection. However, as Boltzmann understood, the microstates (the initial conditions in Γ_{M_2}) that lead to this anti-thermodynamic behavior 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} .

It is helpful in this context to consider the set of all solutions with initial condition in Γ_{M_2} rather than an individual trajectory. 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 future evolution. In other words, nearby solutions of the equations of motion will in general quickly and considerably diverge in phase-space.⁴ Under the Hamiltonian time-evolution, the set of microstates realizing M_2 at the initial time will thus spread all over phase-space (respectively a hypersurface of constant energy, see below) with the overwhelming majority of microstates soon ending up in the equilibrium-region and only a small fraction of special initial configurations evolving into other states of equal or lower entropy.

All in all, the microscopic analysis tells us that it *cannot* be true that *every* non-equilibrium configuration will follow the second law 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. Simply put, a macroscopic system in an “ordered” state will typically evolve into an “unordered” state, because, due to the huge number of microscopic degrees of freedom, there are just vastly many more possibilities to be in disorder than to be in order.

3.2 The measure of typicality

Throughout this argument, the intuitive notions of *almost all* or *extremely special*, that we used synonymously to *typical/atypical*, are understood in terms of the stationary *Liouville-measure*, i.e. in terms of the *phase-space volume* of the set of microstates with the respective property. Stationarity of the measure, as asserted by Liouville’s theorem, means that the volume of

⁴In the case of an ideal gas, this microscopic instability comes only from the reflection of particles on the walls of the box.

a set of microstates is preserved under time-evolution, or, intuitively spoken, that the Hamiltonian flow ϕ_t behaves like an incompressible fluid on phase-space. Formally: $|\phi_t(A)| = |A|$ for all $t \in \mathbb{R}$ and all measurable sets $A \subseteq \Omega$.

This is clearly an essential feature, for it implies that a) the notion of typicality is *timeless*, i.e. a typicality statement does not depend on “what time it is” and b) the Hamiltonian dynamics “care”, so to speak, about the measure of the macro-regions that played such a central role in the argument. More precisely, statement b) can be understood in the sense that the stationary measure, as a measure on *initial conditions* (at an arbitrary initial time), carries over to a well-defined measure on *solution trajectories*, which is such that the “number” of trajectories passing a phase-space region at any given time is proportional to the size of that region.

Now, to be more precise, we would have to take into consideration that for the discussion of a *perfectly isolated* system, only a small subset of phase-space is actually relevant, because the evolution of its microstate is confined to a hypersurface $\Gamma_E \subset \Omega$ of constant energy E in virtue of energy conservation (and this total-energy will, in general, figure into the specification of a system’s macrostate). In this case, the appropriate stationary measure is not the Liouville measure, but the so-called *microcanonical measure*, corresponding to the induced surface-measure on Γ_E . For simplicity, we will usually omit this distinction and simply refer to “phase-space” and the “measure” or “size” of phase-space regions.

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”⁵, 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 $10^{10^{23}} : 1$ in our example).

⁵the term seems to have been introduced by (Frigg 2009).

2. To define or realize a notion of typicality *relative to the current macrostate of the system*, i.e. to allow us to assert that *almost all* initial conditions in a non-equilibrium macro-region Γ_{M_2} will evolve into equilibrium.

Regarding the meaning of “almost all”, one should note that it is 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 $|B| : |\Gamma_{M_2}| \approx |\Gamma_{M_1}| : |\Gamma_{M_2}| \approx 1 : 10^{10^{23}}$.⁶

3.3 Irreversibility

By incorporating into our analysis what is essentially Boltzmann’s answer to Loschmidt’s reversibility objection, we have, in fact, presented the resolution to the issue 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. Let’s take a closer look at how this *prima facie* contradiction is resolved.

To this end, we recall that it was essential to our argument that it always referred to typical 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. Hence, the relevant notion of typicality when discussing convergence to equilibrium from a non-equilibrium macrostate M_2 is defined by the Liouville measure (respec-

⁶Obviously, this number is so vanishingly small, that it doesn’t matter even if there are just a few, a few thousand, or even millions of possible macrostates of lower entropy into which the system could evolve from M_2 . One should note, however, that the relevance of this estimate relies on the (well-founded) assumption that the macrostate of a system changes on larger time-scales than its microscopic state, so that it’s legitimate to think of the macroscopic evolution in discrete time-steps.

tively the microcanonical 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 bad initial conditions in the equilibrium-region Γ_{eq} , initial conditions, that is, 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. (This follows immediately from stationarity of the Liouville measure together with its invariance under the time-reversal operation $T : (q, p) \rightarrow (q, -p)$, reversing the momentum of every particle.) 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} . (An analogous reasoning applies if we compare two non-equilibrium macrostates M_2 and M_1 with substantially different entropy.) It is this fact and this fact alone that establishes the *irreversibility* of a thermodynamic evolution.

Hence, while the issue of macroscopic irreversibility is certainly subtle, mysterious it is not, and we can assure ourselves that our microscopic analysis offers a clear and precise answer to every well-posed question characterizing the phenomenon it is supposed to explain.

Q: *Why do we find that systems in non-equilibrium evolve into equilibrium, although we never see systems in equilibrium evolving into non-equilibrium?*

A: Because the evolution into equilibrium is typical for non-equilibrium configurations, whereas equilibrium configurations evolving into non-equilibrium (except for small fluctuations) are atypical.

Q: *Why are we able to prepare macroscopic systems in states that, once isolated, undergo an (autonomous) evolution of increasing entropy, but not in states that will undergo an (autonomous) evolution of decreasing entropy?*

A: Because the first task only requires the preparation of ordered, i.e. low-entropic, macro-structures (which, remarkably enough, we are very good

at), whereas the second task would require an exceedingly fine-tuned arrangement of a huge number of microscopic degrees of freedom. For instance, a rather crude method such as “pushing a piston” is good enough to prepare a gas in a low-entropy state. Almost every microscopic configuration that the gas could end up in as a result of this procedure is then a microstate that will evolve into equilibrium. However, although there are (in a measure-theoretic sense) just as many equilibrium configurations that would evolve into non-equilibrium, these are very particular and very exceptional ones out of all the configurations that coarse-grain to “a gas in equilibrium”. The preparation of such a state would thus require precise control over roughly 6×10^{23} microscopic degrees of freedom, which is practically impossible.

Q: *Why does our reasoning, which is based on time-symmetric laws and by which we conclude that a system in a low-entropy state will undergo an evolution of increasing entropy not apply to the reversed motion of that very same system?*

A: Because the result of our analysis is a typicality statement and a system that has evolved from a state of lower entropy is *ipso facto* atypical with respect to its evolution in the reversed time-direction.

3.4 Past Hypothesis and the thermodynamic arrow

All in all, we see that it is impossible to understand macroscopic irreversibility without appreciating the fact that the “second law of thermodynamics” is a *typicality statement*, rather than a *necessary* consequence of the microscopic laws or the result of a logical inference about the behavior of any *individual* system (see also the discussion in section 5.). And it is impossible to understand macroscopic irreversibility without appreciating the sense in which non-equilibrium states are *per se* very special, since the thermodynamic irreversibility that we have “derived” from time-symmetric microscopic laws is only a result of the assumption – or preparation, or de facto existence – of these special, i.e. low-entropy, (initial) macrostates. 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 again describes a perfectly time-symmetric situation.

Above we said that every well-posed question about macroscopic irre-

versibility receives a clear and precise answer from Boltzmann’s statistical mechanics. What we should have said, is that we are leaving aside one deep question that is mostly – though not entirely – orthogonal to our discussion. For if we follow our last thought, we see that the typicality account is ultimately shifting the explanatory burden from why it is that systems in non-equilibrium relax into equilibrium, to why it is that we find systems in such special states in the first place. Of course, as long as we are dealing with boxes of gas, or melting ice-cubes, or other confined systems, their low-entropy (initial) states will always be attributable to influences from “outside”, i.e. to the fact that these systems are actually part of some larger system (possibly containing a physicist, or a freezer) before “branching off” to undergo a (more or less) autonomous evolution as (more or less) isolated subsystems. But this presupposes, of course, that these larger systems have been out of equilibrium *themselves*, otherwise they could not have given rise to subsystems with less than 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, much further, in fact, than it would have to be to account for our existence and that of our galaxy. This is what (Goldstein 2001) calls the *hard part of the problem of irreversibility* 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 and the intricacies of the *Past Hypothesis*⁷ stipulating a very “ordered”, i.e. very low-entropy, initial state of our universe. But these questions are far beyond the scope of the present paper and hence shall not be discussed any further (except for a brief remark in section 5, where the “hard problem” is not completely disentangled from the focus of our discussion).

3.5 Recurrence

So far, we have stressed the fact that, as a conclusion of the microscopic analysis, it cannot be true that every non-equilibrium configuration will evolve into equilibrium – only typical ones will. Now, we want to draw the attention to the fact that, as a conclusion of the microscopic analysis, we cannot

⁷The term “Past Hypothesis” is due to (Albert 2000), who also discusses its role in ascertaining the reliability of records and grounding causal asymmetries. The necessity of such an assumption was, however, already noted by Boltzmann in his lectures on gas theory (Boltzmann 1896a, p. 252-253). See also (Feynman 1967) and the pertinent chapters in (Carroll 2010) for a very comprehensible discussion.

claim that a typical system will stay in equilibrium *forever*, but must rather expect that the microstate, evolving in phase-space, will eventually re-enter the non-equilibrium region, corresponding to the system *fluctuating* out of equilibrium into a state of *lower* entropy. The most striking exemplification of this fact is provided by Poincaré’s *recurrence theorem*, stating that a typical solution of the microscopic dynamics will eventually revisit every macro-region (of positive measure) it once passed, meaning, in particular, that a system starting out in a low-entropy macrostate is bound to return to this macrostate some time in the future. But the time-scales on which significant fluctuations out of equilibrium are to be expected are so astronomical that they have no empirical relevance. For the gas-model, for instance, Boltzmann estimates the recurrence time to be about $10^{10^{20}}$ years⁸ – exceeding the age of our universe by many orders of magnitude.

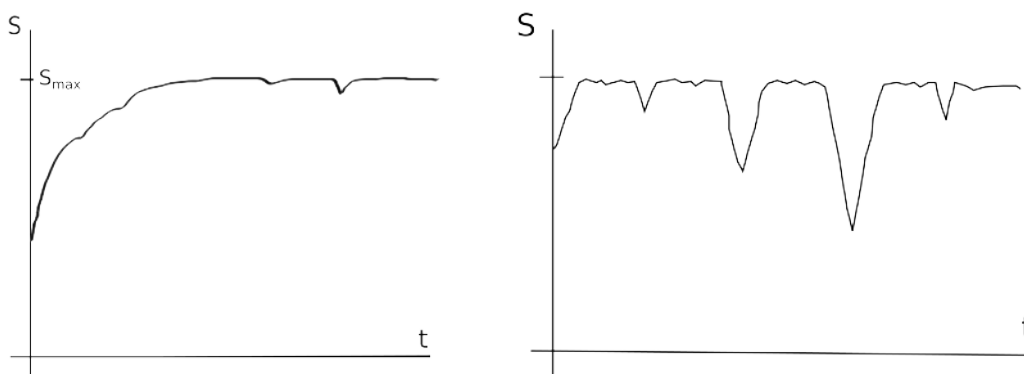


Figure 2: Typical entropy curves on physically relevant time-scales (left) and on time-scales of the Poincaré cycles (right). Note that in the diagram on the right, the proportions are not accurate. The periods of close-to-maximal entropy are really vastly longer than the duration of low-entropy “valleys”.

Whereas we learned from Boltzmann’s typicality account and his reply to Loschmidt’s reversibility objection that the explanation of irreversibility is to an essential degree about *initial conditions*, we learn from his reply to the *recurrence objection* of Zermelo and Poincaré that it is also essentially about *time-scales*. The thermodynamic processes that we observe in our universe and that we want to explain correspond to an entropy-evolution as depicted

⁸See (Boltzmann 1896b) for his illuminating reply to the recurrence objection and his estimate of the recurrence times. Boltzmann does not actually give a numerical value, but only notes that the corresponding number has “many trillion decimal places”.

in the left diagram in Figure 2, that is, systems starting out in a low-entropy state evolving into equilibrium and staying in equilibrium, except for small fluctuations of the entropy about the maximum value. The characteristic time-scale associated with this picture is that of the systems' *relaxation time* (the time it typically takes to reach equilibrium), which may be seconds for the spreading of a gas, minutes for the cooling of a hot bowl of soup, many years for the decay of radon and many billions of years⁹ for the heat death of the universe. But all this is just the blink of an eye compared to the time-scales of the Poincaré cycles for even the fastest reacting system. On *these* time-scales, typical entropy curves will look qualitatively like the graph on the right – which is, by and large, a time-symmetric picture.¹⁰ One will often find discussions of the statistical mechanics and the “second law” being, explicitly or implicitly, concerned with the diagram on the right rather than the diagram the left (see also our discussion of ergodic theory in section 7.) The reader should be aware that these discussions are then hopelessly out of focus, since the time-scales associated with this diagram (i.e. the numbers that we would have to write on the t-axis in any conventional units) are so *ridiculously huge* that they have absolutely no empirical relevance.

4 Proving the second law? Typicality and the H-theorem

Although the formula engraved on Boltzmann's tombstone in Vienna 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 outmost interest in the light of our previous discussion for once, because it illustrates very clearly the need for a typicality argument

⁹probably more like 10^{100} of years, which is about the time it takes for a black hole of several solar masses to evaporate

¹⁰There are as many segments of increasing entropy as there are segments of decreasing entropy, except of course for the initial slope. And the latter appears only if the initial time actually marks the system's (or the universe's) *beginning*, otherwise we should extend the picture analogously towards the past. This picture, pertaining to a universe whose history extends infinitely into the past and into the future, is what Boltzmann considered in his *fluctuation hypothesis* in an attempt to explain the low-entropy state of the universe without assuming a special, low-entropy beginning. Feynman calls this fluctuation hypothesis “ridiculous” (Feynman 1967).

and furthermore because it can be viewed as a concrete implementation of the general scheme that we have presented before. In this context, we have to counter a common misconception, that has most likely arisen from Boltzmann’s first presentation of the H-theorem and persisted despite his more refined argumentation in later writings, namely that the H-theorem and the probabilistic arguments 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)¹¹

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 distinct 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.¹² Understanding these connections, we will also see that many of the objections that are routinely raised against the conclusiveness of the H-theorem are unfounded. To make this case, we shall first review what the H-theorem is actually about and how it is grounded in the microscopic theory.¹³

Recall that the microstate of an N -particle system is represented by a point $X = (q_1, \dots, q_N; p_1, \dots, p_N) \in \Omega$ 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. $X \rightarrow \{(q_1, v_1), \dots, (q_N, v_N)\}$, with $v_i := p_i/m$. The H-theorem is concerned with the evolution of a function $f_X(q, v)$ on μ -space,

¹¹See also the pertinent entry in the Stanford Encyclopedia of Philosophy (Uffink 2008) which presents Boltzmann’s work in statistical mechanics as a series of rather incoherent and ultimately wanting attempts to explain the second law.

¹²Cf. (Goldstein 2012). See also (Goldstein and Lebowitz 2004) for a mathematical analysis corroborating this view.

¹³For a good introduction, see, for instance, (Davies 1977). For a detailed mathematical treatment, see (Spohn 1991), (Villani 2002), (Lebowitz 1981).

sometimes called the *macroscopic profile*, that is supposed to provide an efficient description of the most important (macroscopic) characteristics of the gas in the microstate X . This function is defined as the *empirical distribution* or *coarse-grained density* of points in μ -space. Intuitively, 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 relative 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 go 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 are giving this formula to emphasize that, although $f_X(q, v)$ is technically a *probability measure*, there's absolutely nothing *random* about it. It is more adequate to think of it as a *macroscopic variable*, determined, as it always is, by the microscopic configuration of the system. (More precisely, the usual macroscopic variables of state can be computed from this empirical distribution.) Anyway, what's essential here is that 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 one and 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 is important to distinguish the function $f_{X(t)}(q, v)$ – whose time-evolution is always determined by that of the microstate $X(t)$, i.e. $f_{X(t)}(q, v)$ is always the empirical distribution for the system’s *actual* microscopic configuration – from the solution $f(t, q, v)$ of the Boltzmann equation with initial condition $f(0, q, v) = f_0(q, v)$ ¹⁴. 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*

$$H(f(t, q, v)) := \int f(t, q, v) \log f(t, q, v) dq dv$$

is monotonously decreasing in t .¹⁵

- 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 his 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.

Boltzmann’s derivation of what is now known as the Boltzmann equation is famously based on the *Stoßzahlansatz* or the assumption of *molecular chaos*.¹⁶ This is an assumption about the *relative frequencies* of collisions¹⁷

¹⁴respectively a smooth approximation thereof.

¹⁵Whereas the “true” microscopic H of $f_{X(t)}(q, v)$ will fluctuate and only decrease “on average”.

¹⁶*Assumption*, unfortunately, is not a perfectly accurate translation of the German word *Ansatz*. Whereas the first is sometimes used synonymously with a *premise*, the later has a distinctly pragmatic element and can refer to something more akin to an “approximation” or a “working hypothesis”.

¹⁷The (effective) interaction potential of the particles in the gas is assumed to be very *short-range* compared to their mean free path – this is the meaning of a “diluted” or “low-density” gas. An event where two particles come close enough to each other to interact

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:

$$\mathcal{N}(t, q; v_1, v_2) \propto N^2 f(t, q, v_1) f(t, q, v_2) |v_1 - v_2| dt dq dv_1 dv_2, \quad (2)$$

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. As the mathematically trained reader will readily notice, 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 points, it is true that *if and as long as* the assumption of molecular chaos, respectively equation (2), is valid, statement 1) is correct. Hence, we have to ask: What is the status of molecular chaos and how is it justified?

It is important to keep in mind that there is really nothing *random* about the particle 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. We could then arrange the initial configuration of the system in such a way that “slow” particles will almost exclusively scatter with other “slow” particles and “fast” particles with other “fast” particles (or the other way around). In fact, given that the gas is sufficiently diluted and the range of the pair-interactions sufficiently small, we could even arrange the system in such a way that the particles won’t “meet” at all. But such initial configurations are, obviously, very special ones. For *typical* microscopic configurations is referred to as a “collision”. The potential is, of course, assumed to be repulsive. In the hard sphere model, which is the simplest microscopic model for the Boltzmann equation, a “collision” is quite literally a collision.

figurations, coarse-graining to the initial distribution $f_0(q, v)$, we will 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. Under the constraint that the number of particles in a cell about (q, v) is given by $N \cdot f_0(q, v)$, there are very few possibilities to “arrange” the particles in such a way that (let’s say) each particle meets another one with roughly the same velocity. In comparison, there are vastly many more initial configuration for which the number of particles with velocities v_1 and v_2 colliding near q within an (infinitesimally) short time-interval is (roughly) proportional to $N^2 f_0(q, v_1) f_0(q, v_2) |v_1 - v_2|$. 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.¹⁸

We can 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 $f(t, q, v)$, but it is only for typical initial conditions that the relative frequency of scatterings is actually close to the expectation value. And typical initial conditions are defined, as usually, by the Liouville-measure (respectively the microcanonical measure) restricted to the initial macro-region $\Gamma_I := \{X \in \Gamma_E \mid f_X(q, v) = f_0(q, v)\}$.

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 (2) is approximately 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, because 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 questionable, for the time being, whether a law-or-large-number statement

¹⁸Note, by the way, that there is no issue here as to whether we let the clock run “forwards” or “backwards” – the reasoning is perfectly symmetric with the respect to the time-evolution in both time-directions.

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.

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 affirmative. Given the fact that the microscopic dynamics are highly chaotic, that the number of particles in a gas is huge and the gas, by assumption, very diluted, so that problematic recollisions (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 given by eq. (2)), unless the initial configuration itself was very special. And yet, this is extremely difficult to *prove*, as every mathematician familiar with the problem can testify; 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.¹⁹ However, one should not be confused about the fact that these difficulties are *technical* rather than *conceptual* in nature, and claims to the opposite, that

¹⁹See (Landford 1975) and (King 1975) for the landmark results and (Gallagher et.al. 2012), (Pulvirenti et.al. 2013) for recent extensions to more general interaction potentials. We should note that in these proofs, which all follow the strategy of (Landford 1975), the limiting factor, restricting the validity to a relatively short time-interval, is not, strictly speaking, the propagation of molecular chaos, but rather the existence of the macroscopic (Boltzmann-) dynamics. The two issues are, however, related (for mathematical reasons that we cannot go into here) and which one will appear as the greater obstacle is likely to depend on the strategy of proof.

is, claims to the effect that there is something wrong with the Boltzmann equation – or, for that matter, with Boltzmann’s justification of the equation – are completely unfounded.

In this context, it is also 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 (i.e. as a weak law-of-large-numbers statement), 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 can no longer be adequately described by the Boltzmann equation. It is as astonishing as it is unnecessary that after more than a century, there is still confusion about the question how Boltzmann’s H-theorem squares with the recurrence objection of Zermelo. The answer is simply that the Boltzmann equation is an *effective* description that cannot – and need not – be universally valid, and that, moreover, the time-scales on which the Boltzmann equation is relevant have nothing to do with the time-scales on which Poincaré recurrence is relevant.

All in all, we return to our initial point 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. Although the micro/macro distinction does not appear as prominently in the formulation of the H-theorem, it is essential 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 is the one realized by the *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 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 *Stoßzahlansatz* as the basic assumption in Boltzmann's derivation of his equation, 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 the “most likely” evolutions carry a non-equilibrium distribution into equilibrium *because* almost all microstates are equilibrium states.

Finally, and maybe most importantly, 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 realizing the equilibrium distribution $f_{eq}(q, v)$.

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, Price 2002). This is not correct. There is nothing more time-asymmetric about equation (2) than about the assumption that the probability of drawing a black ball and a white ball from an urn is proportional to the number of black balls and to the number of white balls that this contains as you draw. Of course, it would be hard to see how a time-asymmetric assumption about scattering processes described by reversible microscopic laws could ever be justified, but Boltzmann's arguments contain no questionable ploys like that. 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; this, however, 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*, equation (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, i.e. “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.²⁰

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.

²⁰It may still seem that, since all collisions are reversible, it cannot be the case that the time-reversal of a “good” configuration, for which (2) approximately holds, is a “bad” configuration for which doesn't. This is not correct, though, since the time-reverse of a scattering process $v_1 + v_2 \rightarrow v'_1 + v'_2$ is not $-v_1 + -v_2 \rightarrow -v'_1 + -v'_2$, but $-v'_1 + -v'_2 \rightarrow -v_1 + -v_2$, that is, a reversal of the time-direction leads to completely different “pairings” of incoming velocities in the scattering processes.

5 Typicality and the status of macroscopic laws

5.1 The ‘logic’ of typicality statements

Concerning the philosophical community, 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 reduction²¹ and the related *deductive-nomological models* of physical explanation did not quite capture the subtleties of Boltzmann’s arguments.²² According to these often criticized yet very persistent theories, a microscopic explanation of the second law of thermodynamics – respectively 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 sense in which this characterization is correct, although to get a grip on what this sense is, we will have to say more about what we mean by “derive” and what we mean by a “macroscopic law”. 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 $\forall x(F(x) \Rightarrow G(x))$, 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

²¹For a recent defense of this schemes, see (Dizadji-Bahmani et.al. 2010).

²²Physicists and philosophers who *did* appreciate the relevance of typicality and have written about it include (Maudlin 2007, Bricmont 2001, Dürr 2009, Goldstein 2012, Zanghi 2005). We owe many of the thoughts expressed in this paper to their teachings.

exist initial conditions for which the gas will expand. There are also initial conditions for which the gas will contract. And (possibly) 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 will practically *always* be 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 some fine-tuned arrangement of microscopic degrees of freedom.²³

Note however that the relevant statement is now, logically and syntactically, a proposition about G rather than a proposition about any particular 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.²⁴

5.2 Typicality vs. probability

Now clearly, the more common way of speaking is not to say that a macroscopic feature G (a certain type of macro-evolution, a statistical regularity, a numerical relation between macroscopic variables of state, etc.) is typical, but that $G(x)$ is *very likely* or that we infer $G(x)$ *with high probability*. The tricky thing about this way of speaking, though, is that while almost everybody will consent that it is saying something right and relevant, it is hard to

²³Thanks to Jenann Ismael for this insight.

²⁴We note that a superficial look on the *modus operandi* of statistical mechanics may be deceptive here. Indeed, a quantitative derivation of a thermodynamic law will often be based on a *statistical hypothesis* about the distribution of certain microscopic quantities or events within an individual system (or sometimes an ensemble of systems) – think of Boltzmann’s Stößzahlansatz or the assumption of isotropically distributed particle velocities that enters the derivation of the ideal gas law. However, as we have demonstrated for the Stößzahlansatz in section 4 the “statistical hypothesis” itself is ultimately justified by the fact that the assumed distribution of microscopic quantities is that of a *typical configuration* (respectively a typical ensemble) with respect to the corresponding macrostate. The derivation of a thermodynamic law is then, again, most aptly understood as a result about *typical systems*, rather than a (logical or probabilistic) inference about a *particular* system.

get to people to agree on what exactly that is. Indeed, such a probabilistic statement concerning a “macroscopic law” must 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 this context, but want to make a few general points to capture the intricacy of the issue.

- 1) 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.
- 2) The fact that we are generally ignorant about the 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.
- 3) 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 in the universe that give rise to subsystems (e.g. boxes of gas) in non-equilibrium configurations, than about the entropy-increasing processes going on within these subsystems, once they are suitably isolated and free of macroscopic constraints. 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.

So, 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 once, it is important to note that, 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.²⁵ 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 (1713) called *moral certainty* or *moral impossibility*.²⁶

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” (that is, *information* or *knowledge* or *believe*) about the actual (current or initial) microstate of a system. A typicality statement refers to nothing more and nothing less than the fact that a certain (coarse-grained/macroscopic) property or feature or behavior of a physical system is typical according to the microscopic laws, i.e. that it is the kind of feature or property or behavior that our fundamental theory predicts for an overwhelming 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 temperature will melt. According to the laws of quantum mechanics, a collection of point-particles, shot successively through a double-slit, will typically (though not

²⁵On 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.

²⁶Such typicality statements can be understood in the sense of *Cournot’s principle*, which is one of the basic principles underlying the philosophy of probability 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 about this topic.

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 as a science and a means to understand the world.) 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 this chapter and, in fact, our entire discussion, namely: What is the connection between the macroscopic regularities that we want to account for and the underlying microscopic laws? Another way to put it is to ask: What is the nomological status of the (so called) 'macroscopic' or 'statistical' laws?

5.3 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 (that is, to hold only effectively or "very likely"), but the way in which the empirical regularities expressed by these laws turn out to be *contingent* rather than *necessary* truths about the world that we live in. 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 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 often 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 do occasionally jump up in the air (while cooling off accordingly to account for the conservation of energy) simply because a large number of particles happened to move in the same (upward) 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”.²⁷

And yet, we would insist, it is 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.

In (Kripke 1972), Saul Kripke famously explained the difference between logical and nomological (or natural) 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: the properties of our world that are *typical* for these laws, are facts and regularities 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 our universe to ensure that these properties are realized. However, he barely had to pay any attention to this choice, since almost every initial condition he could have picked (compatible with certain macroscopic constraints he had settled on) would have been fine. For instance, after deciding to create a universe hospitable to the kind of ordered macro-structures that we find in ours, and after deciding to impose on this universe the kind of fundamental laws that we find to hold in ours, God, in principle, still had a choice as to whether or not this universe should also abide by the second law of thermodynamics. However, almost every possible choice of initial conditions (compatible with the Past Hypothesis, see below) would have lead to a universe in which the

²⁷Of 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, but in which the second law of thermodynamics fails to hold in branching systems just so often as to make a fool out of physicists.

second law, as in ours, describes a true statistical regularity – whereas God would have to be utterly malicious to arrange the initial configuration of the universe in such a way that it doesn't.

Typicality, we remark, is a very general way of understanding statistical regularities supervening on deterministic laws. Consider, for instance, the stock example of the coin toss. For all we know, the initial conditions of our universe could have been such that almost every coin ever to be tossed landed on 'heads' rather than 'tails', or such that 'heads' came out twice as often as 'tails', or three times more often, or so on... But such initial conditions are, of course, atypical. Hence, if we believe that there is some deep, explanatory connection between the (rotational) symmetry of the mechanical laws, the symmetry of a balanced coin and the empirical fact that the frequency of 'head' and 'tails' in a sufficiently long series of coin tosses comes out approximately 50:50, this connection is given by *typicality* – at least, we wouldn't know how else.²⁸

Returning to the more intricate issue of the second law of thermodynamics, we have to note one subtlety in connection with the *Past Hypothesis*, the assumed low-entropy initial state of the universe that is supposed to account for the origin of low-entropy structures, the reliability of records, and so on (Albert 2000). According to the Past Hypothesis, the initial *macrostate* of our universe was a very *special* one, marking one end of the thermodynamic arrow of time. Nevertheless, with respect to this low-entropy macrostate, the initial *microscopic configuration* of the universe was *typical*, explaining the increase of entropy in the universe as a whole and in any of its branching subsystems. All in all, there is thus not a contradiction, but a distinct *tension* between the typicality account and the Past Hypothesis and the resolution of this tension is considered by many as one of the most profound problems of modern physics.²⁹

So, 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

²⁸For typicality as a way to understand “deterministic probabilities” in physics, see (Maudlin 2007), (Dürr 2009), (Zanghì 2005)

²⁹See, for instance, the discussion in (Penrose 1999) and his “Weyl curvature hypothesis” as a proposal for an additional law restricting the possible initial states 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 readable and comprehensive discussion of the problem and (Carroll 2010) as well as (Carroll and Chen 2004) for attempt to dispose of the Past Hypothesis altogether.

in the future) is to ask why, in fact, 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 unconvincing. Explanations have to end somewhere. If we can establish that a certain property is typical of a particular kind of system, this should elevate any sense of wonder or mystery or puzzlement that one might have had upon finding such a system instantiating the respective property. Hence, we should consider the phenomenon to be reasonably and conclusively and convincingly *explained* on the basis of the microscopic laws. Similarly, if we can establish that certain macroscopic feature or behavior or regularity is *typical* for a certain kind of system, then we should by all reasonable means *expect* to find this feature or behavior or regularity realized in a given system of 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 we find ourselves in towards the world is necessarily one in which all we can ever hope to know about its physical state is compatible with a plurality of fundamental, i.e. 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.

We shall emphasize once 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 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 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.³⁰

³⁰As was put so nicely by Mathias Frisch (private communication).

5.4 A comment on the ‘measure zero problem’

Despite the many subtleties involved in the concept of typically that would require a much deeper philosophical investigation than we can provide here, 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 about Boltzmann’s statistical mechanics are actually misunderstandings about the “logic” of typicality statements. More concretely, we have found that one of the most common sources of confusion about the typicality account 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; in other words, it approaches equilibrium simply because equilibrium microstates are typical and non-equilibrium microstates are atypical (with respect to Γ_E and μ). [...] This is wrong. If a system is in an atypical microstate (which it is by the Past Hypothesis), 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 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. It does not somehow lie in the ‘nature’ of atypical states to evolve into typical ones. (Frigg 2009, pp. 8–9).

Before we comment on this objection, we have to make the cautionary remark that the way in which we have used the term “typical” throughout this paper

is somewhat different from its use in Roman Frigg’s elaboration. There is, of course, a sense in which it is correct to say that equilibrium states are typical and non-equilibrium states atypical with respect to all possible configurations of the system. The relevant statement would then be that *typical non-equilibrium configurations* will evolve into such a typical state, or that *typical trajectories* will spend most of the time in typical states. In our discussion, we have been careful to use the notion of “typicality” only in the latter sense, referring to initial conditions, respectively solution-trajectories conditioned on a given initial macrostate – a distinction which unfortunately, seems to have gone unappreciated by Frigg, but may have helped to resolve his dissatisfaction.

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 actual micro-evolution carries the microscopic configuration into an equilibrium state. The relevant assertion here is that the regions of space-time that do not correspond to the thermodynamic equilibrium are extremely special. And the claim is then that 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 what this means, in other words, is that typical non-equilibrium states will evolve into equilibrium on relatively short time-scales and that typical equilibrium states will remain in equilibrium (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 Roman Frigg? It is 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 this particular region is a set of measure 1, thus covering almost the entire available phase-space. *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.

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 have just noted) even the entire trajectory of a system will generally constitute a set of measure 0. Although this observation receives ongoing interest from part of the philosophical community and is often presented as a serious objection to typicality arguments (for instance by Frigg himself (p.23), but see also (Sklar 1993)), we don’t think that it causes much of an embarrassment for the reasoning that we have presented.³¹

There are facts and regularities that can be explained on the basis of the fundamental laws of physics 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, yet 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 – there are countless possibilities how any of the objects could be placed slightly differently, or replaced by some other object – 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 seem well-suited for creating confusion where none is due.

6 The role of the dynamics

6.1 The role of mathematics

Statistical mechanics is often taken as the epitome of a (successful) reductionist enterprise in physics. The main goal of such an enterprise is to assure

³¹Thanks to Tim Maudlin for very helpful discussions on this issue.

logical and conceptual consistency between different levels of description of the physical world. To this end, the physical analysis has to tell us how the macroscopic regularities, described by the second law of thermodynamics, can be justified or explained on the basis of the underlying microscopic theory. By which standards are we supposed to evaluate such an account, though? According to the most prominent philosophical position, a microscopic explanation of a macroscopic regularity, respectively a reduction by the microscopic theory, is supposed to 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. And although these deductive-nomological models of explanation (and the related Nagelian schemes of physical reduction) have been regularly criticized and rejected for various reasons, their basic understanding of what it is that a microscopic description has to achieve in order to provide an account of a macroscopic regularity has remained quite pervasive. We don’t mean to get involved in the larger debate taking place in the philosophy of science, to which more competent people have made more elaborate contributions, but make a rather simple point as to why we think this understanding can turn out to be quite debilitating for the issue at hand.

Clearly, a lot hangs on the notion of a “derivation” and one should note that the term can in practice mean something different for a physicist than for a mathematician or a logician, depending on the degree of rigor that it is supposed to imply. If, however, we side with the latter on this issue and understand the term in the strict sense as referring to a *logical deduction* from precisely stated premises or axioms, i.e. a mathematical *proof* about a precisely specified microscopic model, it is just utterly naive to think that this is even within our means when we are dealing with complex macroscopic phenomena involving about 6×10^{23} microscopic degrees of freedom (after all, even simplifying assumptions would ultimately have to be justified or proven on the basis of the microscopic theory). And even in a suitably specified *thermodynamic limit* (where the number N of particles goes to infinity and other quantities in the microscopic model scale accordingly), proving convergence to equilibrium for a more or less realistic microscopic model remains an extremely difficult and largely unresolved problem of mathematical physics. That aside, the mathematical physicist and the physically minded mathematician know very well that not every explanation can be turned into a proof and that not every proof is explanatory. In practice, the “deductive” process that bridges the gap between the simplicity and precision of

the fundamental laws and the intricacy and complexity of the physical world is rarely just a series of precise assumptions and logical inferences – and in fact is not supposed to be. For whereas it lies in the nature of mathematical proof and logical deduction that the truth of the conclusion depends rigidly on the truth of the premises (the former is a “truth-preserving function” of the latter) it is essential for a good physical explanation to be robust against small “perturbations” of the underlying assumptions, which, after all, may themselves arise from a process of approximation and idealization without even the aspiration of factual truth.³²

This realization, which stands somewhat antithetically to the neat meta-scientific theories of the logical positivists, leads to a series of interesting questions, for instance: What can we learn from a mathematical proof applying (let’s say) to a class of “well-behaved” interaction potentials which, however, does not include any of the potentials that we believe to model real-life systems most realistically? What, for instance, do we learn about the thermodynamic behavior of a gas by studying a “hard-sphere model”, since no one actually believes that gas-molecules behave exactly like tiny billiard balls?

To develop an answer to these questions would lead us far beyond the scope of this paper. What we mean to demonstrate, though, is that the relationship between physical understanding and mathematical proof, between explanation and derivation of a (macroscopic) phenomenon is, for a variety of reasons, much more complicated than has often been acknowledged. Anyway, the typicality account that we have presented in this paper is an explanation or an explanatory scheme – not a proof. It tells us what it is that we would have to prove, if a rigorous mathematical theorem is what we are after, but it is neither able, nor intended, nor pretending to shortcut any of the great technical challenges that constitute such an endeavor. Overcoming these challenges and proving the thermodynamic evolution of a suitably complex system is the kind of feat that would get you a Fields Medal (provided you’re young enough, of course). Whether it would give you a better understanding of the second law of thermodynamics is, again, a different question.

³²See (Schwartz 1992) for a beautiful elaboration on this point.

6.2 The role of the dynamics. A reply to critics

The reason for making all these general remarks is that we are irritated by the kind of criticism that the typicality account has been facing from several authors, who have strongly rejected this explanation of the second law of thermodynamics for a lack of mathematical rigor and an alleged failure to make precise the “dynamical assumptions” on which the argument rests. Roman Frigg and Charlotte Werndl even go as far as declaring that the typicality account is “mysterious” because the “connection with the dynamics” is unclear (Frigg and Werndl 2013, p. 918).

Jos Uffink writes on a similar note (as a conclusion to his “counter-example” recited by Frigg and discussed in the previous section):

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

We have both very much and very little to say about these objections. For once, while Uffink is entitled to his epistemic standards, we can only repeat our point that 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 epistemic standards and more a matter of keeping separate things separate is the reference to the reversibility objection that Uffink makes in the same breath. For the reversibility objection, we would 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 1874. Boltzmann gave a perfectly conclusive answer soon after (at least conclusive enough to convince a reasonable person, to borrow Uffink’s terminology) and we now have a very good understanding of how irreversible macroscopic behavior can arise from reversible microscopic dynamics. Producing a rigorous result about the macroscopic behavior of a realistically complex model is a very different issue, involving a very different set of (technical) problems,

and the reversibility objection is actually not one of them. (And even if it was, it is not clear why “an assumption about the dynamics” would be of any help, since one thing that the microscopic dynamics certainly are, by assumption, is reversible.)

All in all, it is not clear what exactly Uffink is confused about: typicality, irreversibility, a lack of mathematical precision, or the role of the dynamics. It just seems to be a lot. Frigg and Werndl, very much to their credit, state more clearly what they have in mind. In (Frigg and Werndl 2013) they explain:

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 (see, for instance, Goldstein 2001). [...] 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 rise to TD-like behaviour possess. [pp. 4–5]

Such demands, however, are also much less reasonable than it might seem. For starters, it is good to keep in mind that we are not actually concerned with dynamical system theory other than as a mathematical tool for where it’s useful, that is, we do not really care about measure-preserving flows on phase space *in general*, but always have in mind a phase-flow generated by a huge number of interacting particles, constituting the kind of macroscopic system that we want to describe. Obviously, whenever we study a particular 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. On the other hand, we understand from Boltzmann’s analysis that 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 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 somewhere 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?

We should also keep in mind that dynamical considerations, to a certain degree, already figure into the partitioning of a system's phase-space into macro-regions and the determination of the volume (respectively the entropy) corresponding to each one of them. In particular, the equilibrium state, while always being the state of maximal entropy, can look very different depending on the broad characteristics of the microscopic interactions. Think, for instance, of the equilibrium state of a mixture of oil and water, which, as we all know, is very different from the homogeneous solution into which a hydrophilic compound will evolve. Or think of a gravitating system, whose thermodynamic evolution is one of clumping and collapsing, very much opposed to that of a (high temperature) gas on which our discussion has focused.³³ But once the stage for the typicality account is thus set, there is just no mystery as to why a system in non-equilibrium should undergo a thermodynamic evolution. Indeed, the dynamics would have to be utterly special to *avoid* carrying (all but “a few”) non-equilibrium configurations into larger and larger macro-regions. It is precisely this 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 simple mathematical prop-

³³The reason being that, for an attractive potential, particles moving closer together pick up kinetic energy, so that, roughly speaking, a macro-configuration that is spatially more concentrated corresponds, at constant energy, to a much larger phase-space volume in the momentum-variables.

erties characterizing the dynamics of roughly 10^{24} interacting particles.³⁴

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.³⁵ In fact, the 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.³⁶ Presumably, what the authors mean to say is that if we considered uncoupled harmonic oscillators as a model for the gas, the argument for its thermodynamic behavior wouldn'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. The only thing that the example actually demonstrates, is that a system of uncoupled harmonic oscillators is not a good model for a gas – which is hardly a surprising discovery.

Now if any of this is a reason to mystify a physical explanation, why stop at thermodynamics. Why not be equally puzzled about the phenomenon of *sound* (let's say), unless we see rigorous mathematical proof of the propagation of sound-waves in a Van-der-Waals gas or provide a precise characterization of all the Hamiltonians leading to such a behavior? Boltzmann's probabilistic explanation of the second law of thermodynamics is not any more *question-begging*, it is just more *subtle* in that it rests, fundamentally, on a typicality reasoning, rather than a mechanistic picture of how microscopic dynamics *produce* a certain microscopic effect.

Indisputably, concerning the microscopic derivation of the second law of

³⁴In section 7 we will explain why the commonplace answer, which is the appeal to ergodic properties that also Frigg and Werndl advocate, doesn't get us anywhere.

³⁵Which, by the way, is neither much better nor much worse than other "counterexamples" we've heard on similar occasions, e.g. that a system will not converge to equilibrium if the Hamiltonian is identically 0, i.e. if the particles don't move at all! Hence, the reason why we are spending so much breath replying to a short parenthesis is that we've found it to be actually quite representative of a common line of argument employed by philosophers of science.

³⁶See also the analysis of a system of anharmonic oscillators in (Bricmont 2001).

thermodynamics, 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.

7 Against (epsilon-)ergodicity

The intellectual attractiveness of a mathematical arguments, as well as the considerable mental labor involved in following it, makes mathematics a powerful tool of intellectual prestidigitation – a glittering deception in which some are entrapped, and some, alas, entrappers. Thus, for instance, the delicious ingenuity of the Birkhoff ergodic theorem has created the general impression that it must play a central role in the foundations of statistical mechanics. [...] The Birkhoff theorem in fact does us the service of establishing its own inability to be more than a questionably relevant superstructure upon [the] hypothesis [of absolute continuity]. (Schwartz 1992)

One of the most tenacious misunderstandings concerning the foundations of statistical mechanics is the crucial role that many authors have ascribed to Boltzmann's ergodic hypothesis, respectively to the concepts of modern ergodic theory.³⁷ Ergodicity, or stronger properties higher up the ergodic hierarchy (see (Berkovitz et.al. 2011)), have thereby been assigned various tasks: to justify the choice of the microcanonical measure as the unique stationary and absolutely continuous measure on the energy-hypersurface,

³⁷While the original ergodic hypothesis didn't even appear anymore in Boltzmann's second "lectures on gas theory", it was later revived, in more modern form, by the groundbreaking work of Birkhoff, von Neumann and Khinchin that established ergodic theory as a remarkably productive (and indeed quite beautiful) field of mathematics, whose physical relevance, however, was – and probably still is – hopelessly exaggerated.

to explain the relevance of Gibbsian ensemble averages by identifying them with time-averages of a single system, or, most to the point of this paper, to explain the convergence from non-equilibrium into equilibrium.

To our knowledge, one of the first philosophical papers to question the explanatory merits of ergodicity was Lawrence Sklar's 1973 publication (Sklar 1973). Sklar entered a much-needed caveat in a time when ergodic theory was one of the hottest topics in mathematical physics. His critique, however, was quite subtle and we would say inappropriately subtle, since the way in which ergodic properties ultimately fail to connect to the relevant physical circumstances and fulfill the explanatory role that had been bestowed on them is actually rather striking, as has been pointed out more clearly in (Schwartz 1992), (Goldstein 2001) or (Bricmont 1995).

Nevertheless, the question that went on to preoccupy people most was whether or not any of the usual gas-models studied by physicists actually *is* an ergodic system, and serious doubts were articulated, for instance, by (Earman and Rédei 1996) (mirroring a similar discussion in the mathematical literature, see for instance (Smale 2000)) that seem to have had some impact. The problem with the question as such is that it inevitably and prematurely reaffirmed its own relevance. Furthermore, the subject remains to some degree a matter of faith, as it is extremely difficult to prove ergodic properties for any realistically complex system.³⁸ Anyway, it is in the light of this debate that (Vranas 1998) and later (Frigg and Werndl 2012, 2013) proposed the weaker (although more artificial) notion of *epsilon-ergodicity* to capture the way in which systems that fail to be ergodic may turn out to be “almost ergodic”.

The remarks that we want to add to this discussion are primarily a comment on their most recent publication (Frigg and Werndl 2013), reasserting the idea of a crucial role of ergodic properties for the explanation of thermodynamic behavior in a self-proclaimed attempt to “demystify” the typicality account. Before we lose ourselves in more general remarks, let's recapitulate what the issue is actually about.

Ergodicity is usually defined in terms of a rather abstract property of dynamical systems (a dynamical system is called ergodic if the invariant sets have measure one or zero), which, however, implies particularly nice behavior for typical solutions. We can call a solution (i.e. a flow-line) of the dynamical system *ergodic* if the proportion of time it spends in any possible region of

³⁸Although it seems to be correct, in general, that sufficiently chaotic systems have good ergodic properties. (Eckmann and Ruelle 1985).

phase-space over its entire history (i.e. in the limit $T \rightarrow \infty$) corresponds to the proportion of phase-space volume occupied by that region. Formally:

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \mathbb{1}_A(X(t)) dt = |A|/|\Omega|, \quad (3)$$

where $\mathbb{1}_A(x)$ is the characteristic function of the set $A \subseteq \Omega$. This is roughly equivalent (although a bit stronger than) the statement that the solution-trajectory covers the phase-space densely, i.e. that it comes arbitrarily close to every single point in phase-space, thus establishing the connection with the original (quasi-)ergodic hypothesis of Boltzmann.³⁹

The relation between “ergodic systems” and “ergodic trajectories” is the following: A dynamical system is ergodic if and only if almost all microstates, that is, all microstates except for a set of measure zero, evolve on an ergodic trajectory. *Epsilon-ergodicity* simply relaxes this condition to “all microstates except for a set of measure $\leq \epsilon$ ” (the relevant measure in both cases is the normalized microcanonical measure on the energy-hypersurface). As mentioned before, it is extremely difficult to assert that any realistic system is actually ergodic. On the other hand, *every* dynamical system is epsilon-ergodic for a sufficiently large value of ϵ , and we understand Frigg and Werndl as saying that there is numerical evidence indicating that typical Hamiltonians in a relevant class of gas-models are epsilon-ergodic for reasonably small values of ϵ that wouldn’t make the property entirely trivial (the two are generally silent about the order of magnitude of ϵ , but we assume in their favor that this is what they mean).

Now how is this supposed to figure in an explanation of thermodynamic behavior? Frigg and Werndl argue that if we assume that a system is epsilon-ergodic (for some reasonably small value of ϵ), we can conclude that typical solutions, in the sense of “all solutions except for a set of initial conditions of measure $\leq \epsilon$ ”, are ergodic. Then, according to the two authors, the story continues as follows:

Consider an initial condition x that lies on an ergodic solution. The dynamics will carry x to [the equilibrium region] Γ_{Meq} and will keep it there most of the time. The system will move out of the equilibrium region every now and then and visit non-equilibrium states. Yet since these are small compared to Γ_{Meq} , it will only

³⁹See for instance (Ehrenfests 1911) on Boltzmann’s ergodic hypothesis, or (Sklar 1973, pp. 77) for a more recent discussion.

*spend a small fraction of time there. Hence the entropy is close to its maximum most of the time and fluctuates away from it only occasionally. Therefore, ergodic solutions behave TD-like [thermodynamic-like].*⁴⁰ (Frigg and Werndl 2013)

One thing that this argument might accomplish is to demonstrate a bit of the appeal of ergodic theory. Whereas (as we saw) it's extremely difficult to prove anything in a rigorous manner just by following the rationale of the combinatorial argument, Frigg and Werndl seem to base their conclusion on a rigorous mathematical statement in form of eq. (7). Unfortunately, though, even the most elegant mathematical expression comes short of being explanatory if it fails to connect to physics. As we will see, the problem with Frigg and Werndl's conclusion, in particular with their claim that an ergodic evolution will carry a non-equilibrium state into equilibrium, is that in the sense in which the statement is correct, it is irrelevant, whereas in the sense in which it *pretends* to be relevant, it's a non-sequitur.

There are various reasons why ergodicity cannot be essential to the explanation of the second law of thermodynamics, the most important being the stark discrepancy between the characteristic time-scales of thermodynamic behavior on the one hand and ergodic behavior on the other. The relevant time-scales for ergodic behavior, the time-scales, that is, on which it starts to matter that trajectories “wind around” the energy hypersurface, are those of the Poincaré cycles i.e. in the order of $10^{10^{20}}$ years!⁴¹ To argue about a system with a macroscopic number of degrees of freedom by means of the ergodic limit is thus to argue about an average over a period of time that is far beyond imagination and even further beyond physical relevance.

It seems to us that the kind of account that Frigg and Werndl promote is motivated, at least in part, by the desire to specify a condition for thermodynamic behavior that can apply to *particular solutions*, i.e. to state an assumption about the dynamical evolution of an *individual* system which implies – and thus supposedly explains – its “thermodynamic-like” behavior. However, it can be readily understood that whether or not the macro-

⁴⁰Frigg and Werndl, alongside other authors, insist on the term “thermodynamic-like behavior” instead of “thermodynamic behavior”, to draw the contrast between the Boltzmann entropy that fluctuates, and the Clausius' formulation of the second law, according to which the entropy of a closed system is never decreasing (Frigg and Werndl 2013, p. 3). We did not adopt this terminology, nor do we find it particularly helpful.

⁴¹Cf. the discussion in section 3.5. This point was, in fact, already made by the Ehrenfests (Ehrenfests 1911, p. 61) about a century ago.

evolution of a system shows the kind of thermodynamic behavior that we want to explain has really nothing to do with whether or not its micro-evolution follows an ergodic solution.

Ergodicity is certainly *not necessary* for thermodynamic behavior since virtually any generic trajectory through phase-space would spend most of the time in the equilibrium-region; To do so, it need not cover phase-space densely anymore than a person's travel routes need to cover the surface of the earth densely to account for him spending most of the time outside the vatican.

And ergodicity is certainly *not sufficient* for thermodynamic behavior, since it is a *time-symmetric* property, that is, the time-reversal of an ergodic solution is also an ergodic solution. But this means, of course, that a derivation of an irreversible macro-behavior cannot possibly be a matter of asserting an ergodic evolution of the microstate.

To put it differently, viz. in the jargon of the previous sections, if the “good” (thermodynamic) solutions are ergodic, then the “bad” (anti-thermodynamic) solutions are ergodic as well. Ergodicity, of course, implies that on the long run – which might be a very very very long run, expressed by the limit $T \rightarrow \infty$ – all of these solutions will *eventually* return to equilibrium and spend overall most of its history, from now to eternity, in that state; however, as emphasized before, this assertion doesn't tell us anything about the evolution of a system on time-scales relevant for explaining the thermodynamic behavior that we observe in nature.

Indeed, if we considered a system in a low entropy state and had good reasons to believe that it's described by an ergodic solution of some Hamiltonian dynamics, it could be an ergodic solution that will rapidly converge to equilibrium, but it could also be an ergodic solution for which the entropy will further decrease – or remain constant – over the next few seconds, or hours, or years, or billions of years...⁴² Of course, it would seem that if our aim was to predict that a gas will either expand or contract or stay roughly the same, we could have turned to classical logic rather than dynamical system theory. The crucial and subtle point is still to argue that one alternative is much more likely than the other, or better, that one alternative is in fact

⁴²The authors briefly acknowledge the issue of “relaxation times” in (Frigg and Werndl 2013), where they go on to argue by citing numerical results about gas-systems that could be ergodic and seem to converge to equilibrium with very short relaxation times, as if that was somehow empirical evidence in their favor. However, no one disputes that ergodic systems will typically converge to equilibrium with relatively short relaxation times. What is disputed is the claim that their ergodicity is essential to this fact.

typical with respect to the system's initial macrostate. This argument was given by Boltzmann and discussed in detail in the first sections of this paper. Ergodicity, we recall, had no part in it.

In addition to the inadequacy of ergodicity for explaining thermodynamic behavior in general, there is a further issue with the notion of epsilon-ergodicity promoted by Frigg and Werndl that makes it a particularly useless concept. That is, even if we established this property for a particular system and some very small value of ϵ , it is very well possible that the order of magnitude of ϵ is comparable to, or even much larger than, the proportion of phase-space (respectively of the energy hypersurface) occupied by a generic low-entropy macrostate (which, after all, may be of the order of $10^{-10^{20}}$ or even much smaller). In this case, epsilon-ergodicity de facto establishes ergodic behavior only for typical systems among those whose initial state is *already* in equilibrium, but doesn't allow us to conclude anything about the more relevant case of systems starting out in non-equilibrium macrostates, since these are potentially among the exception-sets of measure $\leq \epsilon$.⁴³

8 Conclusion

Joel Lebowitz begins his beautiful discussion of Boltzmann's statistical mechanics by noting:

Boltzmann's very original ideas were, perhaps not surprisingly, difficult to grasp for some of his contemporaries. What is surprising is that some of the confusion created by these misunderstandings, the so-called "controversies" with Zermelo and others still persist at present. There is really no excuse for this considering the clarity of Boltzmann's responses and later writings. (Lebowitz 1993, p. 2)

One could add to this last sentence the clarity of Lebowitz' own writings, alongside those of Roger Penrose, Shelly Goldstein, Jean Bricmont, and others. With the present paper, we've made our own effort to eliminate excuses for confusion about Boltzmann's statistical mechanics and his explanation of the "second law".

It is often said, and rightly so, that in science, controversy drives progress. Certainly, when it comes to the foundations or the philosophy of physics, this

⁴³See our discussion in section 3.1. and 3.2 for why Frigg's and Werndl's account fails in this respect and Boltzmann's doesn't.

is as true as ever. And yet, every once in a while, someone just happens to get things right – in which case one can get farther by standing on a giants’ shoulder than by desperately trying to tear him down. And for all that can be said – and that we did say – to translate Boltzmann’s insights into a more modern language, and for all that can be done – and still needs to be done – to supplement his arguments with rigorous mathematical results: Boltzmann just got it right. And his critics got it wrong.

Of course, distinctions have to be made here. The first person who raised the recurrence objection made a very relevant and instructive point. The hundredth’s person to raise the same objection more than a century later, still insisting that it reveals a fundamental inconsistency in Boltzmann’s arguments, is no longer advancing the issue. On the other hand, some of the objections raised by modern philosophers of science, dwelling on the trivial observation that the typicality account is not a rigorous mathematical proof, were rather unrewarding even the first time.

This is not to say that there are no open issues – quite the contrary. Standing on solid foundations, the substantial questions and the most promising paths to move forward become only more clear. In our paper, we have identified several such questions and shown how they can be understood and approached on the basis of Boltzmannian statistical mechanics. From a physical perspective, the deepest question, arguably, concerns the Past Hypothesis, i.e. the origin of irreversibility and the thermodynamic arrow of time in our universe, that appears to lie in a very special macrostate of the early universe. In mathematics, epochal problems like the validity and the solution theory of the Boltzmann equation are only the most prominent examples illustrating the need to develop new techniques to handle the complexity of many-particle physics and the micro-to-macro transition.⁴⁴

Finally, we think that the philosophical debate would greatly benefit from appreciating and further elaborating typicality as a form of non-deductive reasoning and maybe even as a fundamental philosophical category. In this paper, we have tried to demonstrate the great relevance of the concept of typicality for understanding probabilities in physics, describing macro-to-micro reduction and grounding the nomological status of macroscopic laws. And yet, there seems to be disappointingly little precedence in the philosophical literature, so that much remains to be said, more insights to gain and more benefits to reap. However not here.

⁴⁴See (Villani 2002) for a survey of recent developments.

References

- [Albert 2000] Albert, D.: *Time and Chance*. Harvard University Press, Cambridge, London (2000).
- [Berkovitz et.al. 2011] Berkovitz, J., Kronz, F. and Frigg, R.: ‘The Ergodic Hierarchy’, in: Zalta, E. (ed.): *The Stanford Encyclopedia of Philosophy* (2011).
- [Bernoulli 1713] Bernoulli, J.: *Ars conjectandi, opus posthumum. Accedit Tractatus de seriebus infinitis, et epistola gallic/’e scripta de ludo pilae reticularis*, Thurneysen Brothers, Basel (1713). Cited after: Bernoulli, J.: *The art of conjecturing*. The John Hopkins University Press, Baltimore (2006).
- [Boltzmann 1896a] Boltzmann, L.: *Vorlesungen über Gastheorie*. Verlag v. J. A. Barth, Leipzig (Nabu Public Domain Reprints) (1896).
- [Boltzmann 1896b] Boltzmann, L.: Entgegnung auf die wärmetheoretischen Betrachtungen des Hrn. E. Zermelo, in: *Wied. Ann.* 57 (1896), pp. 773-784.
- [Bricmont 1995] Bricmont, J.: Science of Chaos or Chaos in Science? In: *Physicalia Magazine*, 17 (1995), pp. 159-208.
- [Bricmont 2001] Bricmont, J.: ‘Bayes, Boltzmann and Bohm: Probabilities in Physics’, in: Bricmont, J., Dürr, D., et.al.: *Chance in Physics. Foundations and Perspectives*. Springer-Verlag, Berlin, Heidelberg (2001), pp. 3-21.
- [Callender 2004] Callender, C.: ‘There is No Puzzle about the Low Entropy Past’, in: Hitchcock, C. (ed.): *Contemporary Debates in Philosophy of Science*. Blackwell, London (2004), pp. 240-255.
- [Carroll and Chen 2004] Carroll, S. and Chen, J.: ‘Spontaneous Inflation and the Origin of the Arrow of Time’. ArXiv: hep-th/0410270 (2004).
- [Carroll 2010] Carroll, S.: *From Eternity to Here. The Quest for the Ultimate Theory of Time*. Dutton, Penguin Group Inc, USA (2010).
- [Davies 1977] Davies, P.C.W.: *The Physics of Time Asymmetry*. University of California Press, Berkeley and Los Angeles (1977).

- [Dizadji-Bahmani et.al. 2010] Dizadji-Bahmani, F., Frigg, R., Hartmann, S.: Who's Afraid of Nagelian Reduction? In: *Erkenntnis*, 73 (2010), pp. 393-412.
- [Dürr 2009] Dürr, D.: *Bohmian Mechanics*. Springer-Verlag, Berlin, Heidelberg (2009).
- [Earman and Rédei 1996] Earman, J. and Rédei, M.: Why ergodic theory does not explain the success of equilibrium statistical mechanics, in: *The British Journal for the Philosophy of Science*, 47 (1996), pp. 63-78.
- [Eckmann and Ruelle 1985] Eckmann, J.-P. and Ruelle, D.: Ergodic theory of chaos and strange attractors, in: *Reviews of Modern Physics*, Vol. 57, No. 3, Part I (1985), pp. 617-656.
- [Ehrenfests 1911] Ehrenfest, P. und T.: 'Begriffliche Grundlagen der statistischen Auffassung in der Mechanik', in: *Enzyklopädie der mathematischen Wissenschaften mit Einschluss ihrer Anwendungen*, Band 4 (1911), pp. 3-90.
- [Feynman 1967] Feynman, R.: *The Character of Physical Law*. The M.I.T. Press, Cambridge (1967).
- [Frigg 2009] Frigg, R.: Typicality and the Approach to Equilibrium in Boltzmannian Statistical Mechanics, in: *Philosophy of Science*, 76 (2009), pp. 997-1008.
- [Frigg 2011] Frigg, R.: 'Why Typicality Does Not Explain the Approach to Equilibrium', in: Suárez, M. (ed.): *Probabilities, Causes and Propensities in Physics*. Synthese Library (347). Springer, Dordrecht, Netherlands (2011), pp. 77-93.
- [Frigg and Werndl 2013] Frigg, R. and Werndl, C.: Explaining Thermodynamic-Like Behaviour In Terms of Epsilon-Ergodicity, in: *Philosophy of Science*, 78 (2013), pp. 628-652 .
- [Frigg and Werndl 2012] Frigg, R. and Werndl, C.: Demystifying Typicality, in: *Philosophy of Science*, 79 (2012), pp. 917-929.
- [Gallagher et.al. 2012] Gallagher, I., Saint Raymond, L. and Texier, B.: 'From Newton to Boltzmann: the case of short-range potentials'. ArXiv 1208.5753v1 [math.AP] (2012).

- [Goldstein 2001] Goldstein, S.: ‘Boltzmann’s Approach to Statistical Mechanics’, in: Bricmont, J., Dürr, D., et.al.: *Chance in Physics. Foundations and Perspectives*. Springer-Verlag, Berlin, Heidelberg (2001), pp. 39-54.
- [Goldstein 2012] Goldstein, S.: ‘Typicality and Notions of Probability in Physics’, in: Ben-Menahem, Y. and Hemmo, M. (eds.): *Probability in Physics*. The Frontiers Collection, Springer-Verlag, Berlin, Heidelberg (2012), pp. 59-71.
- [Goldstein and Lebowitz 2004] Goldstein, S. and Lebowitz, J.: On the (Boltzmann) entropy of non-equilibrium systems. In: *Physica D: Non-linear Phenomena*, Vol. 193, I. 1-4 (2004), pp. 53-66.
- [King 1975] King, F.: *BBGKY hierarchy for positive potentials*. Ph.D. Thesis, Department of Mathematics, University of California at Berkeley (1975).
- [Kripke 1972] Kripke, S.: ‘Naming and necessity’, in: Harman, G. and Davidson, D. (eds.): *The Semantics of Natural Language* (1972), pp. 254-355.
- [Landford 1975] Lanford, O.E.: ‘Time Evolution of Large Classical Systems’, in: Moser, J. (ed.): *Lecture Notes in Physics*, Vol. 38, Springer-Verlag, Berlin, Heidelberg (1975), pp. 1-111.
- [Lebowitz 1981] Lebowitz, J.: Microscopic dynamics and macroscopic laws, in: *Annals New York Academy of Sciences* (1981), pp. 220-233.
- [Lebowitz 1993] Lebowitz, J.: Macroscopic laws, microscopic dynamics, time’s arrow and Boltzmann’s entropy, in: *Physica A* 194 (1993), pp. 1-27.
- [Maudlin 2007] Maudlin, T.: What could be objective about probabilities? In: *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, Volume 38, Issue 2 (2007), pp. 275-291.
- [Penrose 1999] Penrose, R.: *The Emperor’s New Mind*. Oxford University Press, Oxford (1999).

- [Price 1996] Price, H.: *Time's Arrow & Archimedes' Point. New Directions for the Physics of Time*. Oxford University Press, New York, Oxford (1996).
- [Price 2002] Price, H.: 'Burbury's Last Case: The Mystery of the Entropic Arrow', in: Callender, C. (ed.): *Time, Reality & Experience*. Cambridge University Press, Cambridge (2002), pp. 19-56.
- [Pulvirenti et.al. 2013] Pulvirenti, M., Saffiro, C. and Simonella, S.: 'On the validity of the Boltzmann equation for short range potentials'. ArXiv: 1301.2514v1 [math-ph] (2013).
- [Schwartz 1992] Schwartz, J.: 'The pernicious influence of mathematics on science', in: Kac, M., Rota, G.-C. and Schwartz, J.: *Discrete Thoughts*. Birkhäuser, Boston (1992), pp. 19-25.
- [Shafer and Volk 2006] Shafer, G. and Volk, V.: The Sources of Kolmogorov's Grundbegriffe, in: *Statistical Science*, Vol. 21, No. 1 (2006), pp. 70-98.
- [Sklar 1993] Sklar, L.: *Physics and chance. Philosophical issues in the foundations of statistical mechanics*. Cambridge University Press, Cambridge (1993).
- [Sklar 1973] Sklar, L.: Statistical Explanation and Ergodic Theory, in: *Philosophy of Science*, Vol. 40, No. 2 (1973), pp. 194-212.
- [Smale 2000] Smale, S.: 'On the Problem of Reviving the Ergodic Hypothesis of Boltzmann and Birkhoff', in: Cucker, F. and Wong, R. (eds.): *The Collected Papers of Stephen Smale*. Volume 2. Singapore University Press (2000), pp. 823-830.
- [Spohn 1991] Spohn, H, *Large Scale Dynamics of Interacting Particles*. Springer-Verlag, Berlin, Heidelberg (1991).
- [Uffink 2007] Uffink, J.: 'Compendium of the foundations of classical statistical physics', in: Butterfield, J. and Earman, J. (eds.): *Handbook for the Philosophy of Physics*. Amsterdam Elsevier (2007), pp. 923-1047.
- [Uffink 2008] Uffink, J.: 'Boltzmann's Work in Statistical Physics', in: Zalta, E. (ed.): *The Stanford Encyclopedia of Philosophy* (2008).

- [Villani 2002] Villani, C.: ‘A review of mathematical topics in collisional kinetic theory’, in: Friedlander, S. and Serre, D. (eds.): *Handbook of mathematical fluid dynamics*, Vol. 1. Elsevier Science, Amsterdam (2002) , pp. 71-305.
- [Vranas 1998] Vranas, P.: Epsilon-Ergodicity and the Success of Equilibrium Statistical Mechanics, in: *Philosophy of Science*, Vol. 65, No.4 (1998), pp. 688-708.
- [Zanghì 2005] Zanghì, N.: ‘I fondamenti concettuali dell’approccio statistico in fisica’, in: Allori, V., Dorato, M., Laudisa, F. and Zanghi, N. (eds.): *La Natura Delle Cose. Introduzione ai Fundamenti e alla Filosofia della Fisica*. Carocci, Roma (2005), pp. 139-228.