

Chapter 3

Typicality and the Role of the Lebesgue Measure in Statistical Mechanics

Itamar Pitowsky

Abstract Consider a finite collection of marbles. The statement “half the marbles are white” is about counting and not about the probability of drawing a white marble from the collection. The question is whether non-probabilistic counting notions such as *half*, or *vast majority* can make sense, and preserve their meaning when extended to the realm of the continuum. In this paper we argue that the Lebesgue measure provides the proper non-probabilistic extension, which is in a sense uniquely forced, and is as natural as the extension of the concept of cardinal number to infinite sets by Cantor. To accomplish this a different way of constructing the Lebesgue measure is applied.

One important example of a non-probabilistic counting concept is typicality, introduced into statistical physics to explain the approach to equilibrium. A typical property is shared by a vast majority of cases. Typicality is not probabilistic, at least in the sense that it is robust and not dependent on any precise assumptions about the probability distribution. A few dynamical assumptions together with the extended counting concepts do explain the approach to equilibrium. The explanation though is a weak one, and in itself allows for no specific predictions about the behavior of a system within a reasonably bounded time interval.

It is also argued that typicality is too weak a concept and one should stick with the fully fledged Lebesgue measure. We show that typicality is not a logically closed concept. For example, knowing that two ideally infinite data sequences are typical does not guarantee that they make a *typical pair* of sequences whose correlation is well defined. Thus, to explain basic statistical regularities we need an independent concept of typical pair, which cannot be defined without going back to a construction of the Lebesgue measure on the set of pairs. To prevent this and other problems we should hold on to the Lebesgue measure itself as the basic construction.

This chapter was written by Itamar Pitowsky for this volume shortly before his death. As we did not have Itamar's LaTeX file, it had to be retyped, and proofread by us. Remaining typos are, therefore, our fault. We thank William Demopoulos for reading the chapter and streamlining some of its formulations.

I. Pitowsky (✉)

Edelstein Center, The Hebrew University, Jerusalem, Israel

3.1 Introduction

Consider a finite but large collection of marbles. When one says that a vast majority of the marbles are white one usually means that all the marbles except possibly very few are white. And when one says that half the members are white, one makes a statement about counting, and not about the probability of drawing a white marble from the collection. The question is whether non-probabilistic notions such as *vast majority* or *half* can make sense, and preserve their meaning when extended to the realm of the continuum, especially when the elements of the collection are the possible initial conditions of a large physical system.

A major purpose of this paper is to argue that the task of expanding combinatorial counting concepts to the continuum can be accomplished. In the third section we shall see that counting concepts, which have a straight-forward meaning in the finite realm, also have an extension in the construction of the Lebesgue measure. Moreover, we shall argue that the extension is in a sense uniquely forced, as the famous extension of the concept of cardinal number to infinite sets by Cantor. To accomplish this task a different route to the construction of the Lebesgue measure is taken.

All this relates to the notion of *typicality* [1–9], introduced to statistical physics to explain the approach to equilibrium of thermodynamic systems. This concept has at least three different definitions [9], all entail that a typical property is shared by a vast majority of cases, or almost all cases. Typicality is not a probabilistic concept, this is maintained explicitly [3, 7, 8] or implied, at least in the sense that typicality is robust and “not dependent on any precise assumptions” about the probability distribution [2]. A recent example ([8], page 9):

“When employing the method of appeal to typicality, one usually uses the language of probability theory. When we do so we do not mean to imply that any of the objects considered is random in reality. What we mean is that certain sets (of wave functions, of orthonormal bases, etc.) have certain sizes (e.g., close to 1) in terms of certain natural measures of size. That is, we describe the behavior that is typical of wave functions, orthonormal bases, etc. However, since the mathematics is equivalent to that of probability theory, it is convenient to adopt that language. For this reason, we do not mean, when using a normalized measure μ , to make an “assumption” of a priori probabilities,” even if we use the word “probability”.

However, none of the above papers explain in a precise manner why the Lebesgue measure is a “natural measure of size”, or what is the connection between the continuum notions of “vast majority of cases” or “typical cases”, and the equivalent finite notions which are based on simple counting.

A few modest dynamical assumptions combined with the combinatorial notions do explain the approach to equilibrium. I shall argue that the explanation is a weak one, and in itself allows for no specific predictions about the behavior of the system within a reasonably bounded time interval. Whenever predictions of that kind are made some additional knowledge about the initial condition or dynamics should be added. This is where probability enters the picture. We shall argue this for a finite system in the next section, and consider the infinite case in the 4th section.

Typicality, however, is too weak a concept and it is argued in the last section that one should stick with the full-fledged Lebesgue measure. Typicality does not quite cover measurable subsets whose measure is strictly between zero and one, which we might use in statistical mechanics. Even more seriously, the concept is not logically closed. For example, consider Galton's Board which is a central example in [3, 7]. Knowing that two ideally infinite sequences are typical does not guarantee that they make a *typical pair* of sequences whose correlation is well defined and equal to 0.25. Therefore, the concept of typical sequence cannot be used to explain basic long term statistical regularities. For this we need an independent concept of typical pair, which cannot be defined without going back to a construction of the Lebesgue measure on the set of pairs. Similar observations apply to triples, quadruples, and all k -tuples; in each case typicality cannot be defined just on the basis of the former notions.

3.2 Divine Comedy- the Movie

Consider the set of all possible square arrangements of $1,000 \times 1,000$ black and white pixels. There are 2^{10^6} such arrangements, we shall call each one a picture, and the set of all pictures is our phase space. Imagine that upon his arrival in Hell, a lesser sinner is seated in a movie theater (no air conditioning). The show consists of the following movie:

1. Pictures are projected on the screen at a constant pace of 25 frames a second.
2. The sequence is deterministic, the director has arranged that each picture gives rise to a unique successor. We can assume that the dynamical rule is internal, so that each picture, apart from the first, depends uniquely on the pixel arrangement of its predecessors.
3. The movie goes through all 2^{10^6} pictures, and then starts again. So the show is periodic, but the period is extremely long, more than 10^{301020} years long (compared with the age of the universe which is less than 10^{11} years). The phase space contains all the pictures that were ever shot and will ever be shot, including photo copies of written texts and frames from movies, provided they are cast in the format of a thousand by a thousand black and white pixels. Despite this, the set of pictures that look remotely like regular photographs is very small compared with the totality of pictures. Worse, the set of pictures that contain a large patch of black (or white) pixels is very small. These are just combinatorial facts: the overwhelming majority of pictures look gray, approximately half black and half white, with the black and the white pixels well mixed. The number of pictures with a single color patch of size m decreases exponentially with m .

The conjunction of the three dynamical rules for the movie with this combinatorial observation explains why, in the long run, the movie is extremely boring and

looks gray. It also explains why, in the long run, the frequency of the pictures that have more black than white pixels is (a little less than) 0.5.¹ We have to be clear about the meaning of “the long run” here. In the absence of any detail about the dynamics other than rules **1**, **2**, **3**, we cannot really say how long the long run is. It may be the case that the movie begins with a 50,000 year-long stretch of cinematic masterpieces. However, this cannot last much longer and the movie then settles into almost uniform gray for a vast length of time. Likewise, it is also possible that the director has chosen a dynamics that puts all the pictures with more black than white pixels at the end of the movie. In this case the long run may be very long indeed.

Another way of looking at the long run is to notice that, given the nature of the theater, different spectators arrive at different times. The first picture each newcomer encounters upon arrival can be taken as an “initial condition”. So the answer to the question “how long will it take for the movie to settle into almost uniform gray?” depends on the initial condition. Similarly, the number of frames it takes the time average of an “observable” (a function $f : \{0, 1\}^{10^6} \rightarrow \mathbf{R}$) to stabilize depends on the initial condition.

So far nothing has been said about probabilities, it is clear that the frequencies are just proportions in a finite set. The explanation for the frequencies is straightforward and involves no probabilities. However, the questions that can be answered are limited. On the basis of the three dynamical rules and counting alone we can make no specific forecasts. In the best case we obtain a simple theory which is consistent with what we see.

Probabilistic considerations enter when definite predictions are made, beyond the long run explanations. Given the deterministic nature of the system, probability in this context is invariably epistemic. Consider the claim that the picture to be projected two minutes from now will have more black than white pixels. We can imagine two extreme reactions: A savant spectator (Laplace’s demon) may have figured out what the dynamics is, and knowing the present condition, may calculate the pattern of pixels two minutes from now. The probability he assigns his result is one, or very near one allowing for a possible mistake. At the other extreme, where most spectators are, no information beyond the dynamical rules is available. In this case a natural choice of a prior is the uniform distribution, that is, the counting measure represents the probability.² The probability assigned to the event is thus (slightly less) than 0.5. It is easy to invent stories where partial information is available, with the consequence that the probability can be anything between zero and one.

Now imagine that upon their arrival in Hell heavier sinners are made to watch a different show. They are seated in front of a large transparent insulated container

¹Note that the number of white pixels in a picture may be considered a “macroscopic” observable, whose measurement requires no detailed knowledge of the pixel distribution. If we assume that white pixels emit light, and black pixels do not, we just measure the light emitted from a picture, and compare it with the all white picture.

²This is not a very smart prior, though. It assumes independence, and therefore blocks the possibility of learning from experience.

full of gas at a constant temperature and sulk at it. Nothing much happens of course, and the question is whether we can explain why this is the case on grounds that are similar to the movie story. Here a single picture is analogous to one microscopic state, and the movie as a whole to the continuous trajectory of the microstate in phase space. However, since there is a continuum of microstates it is not clear how to expand the finite concepts to the continuum. In particular, it is not clear what is the meaning of overwhelming majority of microstates, or typical states, or half the microstates, unlike the finite case where we just use the terms with their ordinary meaning. The translation of the dynamical rules **1**, **2**, **3** to the motion of particles is not obvious either.

Boltzmann had a long and complicated struggle with these issues [10]. In some writings he was clearly attempting to associate combinatorial intuition, finite in origin, with continuous classical dynamics. However, he lacked the appropriate mathematics which had not yet been invented, or at any rate, was not yet widely known among physicists. By the time it became available combinatorial and probabilistic consideration were hopelessly mixed up. The idea of typicality goes a long way to disentangle the two issues.

Putting the dynamical questions aside for a while, the next section is devoted to the extension of the relevant combinatorial concepts to the domain of the continuum. It is therefore a chapter in the philosophy of mathematics.

3.3 The Road Less Travelled to Lebesgue Measure

Our purpose is to extend concepts such as *majority of cases*, or *one quarter of the cases*, from the finite realm, where their meaning is obvious, to the domain of the continuum. Extensions of mathematical concepts from one realm to a larger domain that contains it are not necessarily unique, and may result in a large variety of quite different creatures [11]. However, in some cases there are very compelling arguments why one particular possible extension is the correct choice, the most important example being Cantor's definition of the cardinality of infinite sets. I shall argue below that the Lebesgue measure plays a similar role in the extension of combinatorial counting concepts.

Usually the Lebesgue measure is introduced as part of the modern theory of integration, the extension of the definition of the integral beyond the limitations of Riemann's construction. This is consistent with the historical development, and answers the requirements of the mathematics curriculum. Here we take another approach altogether. First note that without loss of generality our efforts can concentrate on the interval $[0, 1]$ with the Lebesgue measure on it. The reason is that every (normalized) Lebesgue space is isomorphic to this space, meaning that there is a measure preserving isomorphism between the two spaces.³ Second, note

³This general result is due to Caratheodory, see [12, page 16].

that the interval $[0, 1]$ can be replaced with the set of all infinite sequences of zeros and ones $\{0, 1\}^\omega$, when we identify each infinite zero-one sequence $\mathbf{a} = (a_1, a_2, a_3, \dots)$ with a binary development of a number in $[0, 1]$, that is, $\mathbf{a} \rightarrow \sum_{j=1}^\infty a_j 2^{-j}$. This map is not 1-1, but fails to be 1-1 only on the countable set of rational numbers whose denominator is a power of 2 (dynamic numbers), hence a set of measure zero. In sum, our construction of the Lebesgue measure is developed without loss of generality as an extension from sets of finite 0-1 sequences to subsets of $\{0, 1\}^\omega$.

We start with the finite case, where the movie of the previous section is the example we want to generalize. We can represent the movie as the set of sequences of zeros and ones of length one million $\{0, 1\}^{10^6}$, where each picture is an element of that set. Consider more generally the set $\{0, 1\}^n$ where n is any natural number, and $A \subseteq \{0, 1\}^n$. Then the *measure* μ_n of A is defined to be

$$\mu_n(A) = 2^{-n}|A|, \tag{3.1}$$

where $|A|$ is the number of elements of A . So that, for example, if $\mu_n(A) = 0.5$ we can say that half the sequences of $\{0, 1\}^n$ belong to A . The size measure has an important invariance property: If $m > n$ then $\{0, 1\}^m = \{0, 1\}^n \times \{0, 1\}^{m-n}$, we can embed every $A \subseteq \{0, 1\}^n$ in $\{0, 1\}^m$ by the map

$$A \subseteq \{0, 1\}^n \rightarrow A' = A \times \{0, 1\}^{m-n} \subseteq \{0, 1\}^m, \tag{3.2}$$

so that $\mu_n(A) = \mu_m(A')$.

With these notations we can formulate the claim made in the movie story, that the overwhelming majority of pictures are approximately half black and half white. Given a sequence $\mathbf{a} = (a_1, a_2, \dots, a_n) \in \{0, 1\}^n$, let $S_n(\mathbf{a}) = \sum_{j=1}^n a_j$ be the sum of the elements of \mathbf{a} , and thus the average number of ones in the sequence is $n^{-1}S_n(\mathbf{a}) = n^{-1} \sum_{j=1}^n a_j$. Therefore, the claim is that for a sufficiently large n the vast majority of sequences satisfy $n^{-1} S_n(\mathbf{a}) \sim 0.5$. Indeed, the weak law of large numbers (LLN) states: *For every $\varepsilon > 0$*

$$\mu_n \left\{ \mathbf{a} \in \{0, 1\}^n; \frac{1}{2} - \varepsilon \leq n^{-1} S_n(\mathbf{a}) \leq \frac{1}{2} + \varepsilon \right\} > 1 - \frac{1}{4n^2\varepsilon^4}, \tag{3.3}$$

so that the left hand side tends to 1 as $n \rightarrow \infty$.⁴

Students usually encounter this or similar finite versions of LLN in a course on probability and statistics. In rare cases the teachers make it a point to distinguish the

⁴The rate of convergence on the right hand side of (3.3) is better than the historical one derived by Bernoulli. We need the stronger result (essentially due to Borel) for later purposes. See [13], page 40.

two meanings of LLN. First there is the familiar one of probability theory concerning, for example, Bernoulli trials with probabilities p and $q = 1 - p$ for the two outcomes. In case the distribution is uniform, $p = q = 0.5$, a formula like (3.3) obtains. The second meaning, the one used here, concerns counting the number of elements in the set between the braces in (3.3), or equivalently, calculating the proportion of such elements in the set of all 0–1 sequences of length n . This combinatorial meaning is much simpler, and is qualitatively apparent by looking at Pascal’s Triangle.

The difference between the two meanings of LLN can be better understood when we consider the conditions for their applications. In the probabilistic case we have to describe by which process the digits in the sequence are chosen, for example, by coin tosses with probability p for “heads”. Subsequently, we have to justify the assumption that coin flips are independent, and finally to explain that LLN is saying that the probability that the average of “heads” lies close to p is large. By contrast, in the application of the combinatorial theorem there is nothing to explain, the process of counting requires no further analysis. As noted, the distinction between the two meanings of the weak LLN is rarely taught in the class-room or mentioned in text books. Moreover, this distinction is never mentioned at all when it comes to the strong LLN, despite the fact that the strong LLN is a *consequence* of inequality (3.3) and σ -additivity (see below).

Moving to the infinite case, consider the set of all infinite 0–1 sequences $\{0, 1\}^\omega$. Given a finite set $A \subseteq \{0, 1\}^n$ we can embed it as a subset of $\{0, 1\}^\omega$ using the same method in (3.2) namely

$$A \subseteq \{0, 1\}^n \rightarrow F = A \times \{0, 1\} \times \{0, 1\} \times \dots \subseteq \{0, 1\}^\omega. \quad (3.4)$$

Call every subset of $\{0, 1\}^\omega$ that has the form of F in (3.4) finite. Summarizing, $F \subseteq \{0, 1\}^\omega$ is finite if it has the form $F = A \times \{0, 1\} \times \{0, 1\} \times \dots$, with $A \subseteq \{0, 1\}^n$ for some natural number n . Of course F has infinitely many elements, but this does not cause confusion as long as the context is clear. Now, define the measure μ of F to be

$$\mu(F) = \mu_n(A) = 2^{-n}|A|. \quad (3.5)$$

As long as only finite subsets of $\{0, 1\}^\omega$ are considered no real expansion of the concept of measure is achieved. Note that the family of all finite subsets is a Boolean algebra, it is closed under complementation and (finite) unions and intersections. The minimal expansion to infinity is achieved by considering countable infinite unions and intersections. Denote the Boolean algebra of finite subsets of $\{0, 1\}^n$ by \mathcal{F} . In other words, $F \in \mathcal{F}$ if F has the form $F = A \times \{0, 1\} \times \{0, 1\} \times \dots$ with $A \subseteq \{0, 1\}^n$ for some natural number n . The σ -algebra \mathcal{B} of Borel subsets of $\{0, 1\}^\omega$ is defined to be the minimal σ -algebra that contains \mathcal{F} . This means that \mathcal{B} is the minimal family of subsets of $\{0, 1\}^\omega$ which contains \mathcal{F} , and is closed under complementation, and under countable unions and countable

intersections of its own elements, to generate \mathcal{B} , one takes countable unions of finite sets, then countable intersections of the resulting sets, and so on.⁵

The measure μ is extended from \mathcal{F} to \mathcal{B} using the σ -additivity rule: If $E_1, E_2, \dots, E_j, \dots \in \mathcal{B}$ is a sequence subsets, disjoint in pairs, i.e., $E_i \cap E_j = \emptyset$ for $i \neq j$, then

$$\mu\left(\bigcup_{j=1}^{\infty} E_j\right) = \sum_{j=1}^{\infty} \mu(E_j). \quad (3.6)$$

Usually, one additional “small” step is taken to complete the construction: Given any Borel set $B \in \mathcal{B}$ such that $\mu(B) = 0$ add every such subset of B to the Borel algebra \mathcal{B} . The larger σ -algebra which is generated after this addition is the Lebesgue algebra \mathcal{L} . The measure μ , which is extended to \mathcal{L} in an obvious way, is the Lebesgue measure.⁶

Why is μ the correct expansion to infinity of the size measure in the finite case? Obviously, the crucial steps in the expansion are the construction of the σ -algebra and the application of σ -additivity. As a consequence new theorems can be formulated and proved, for example, the strong law of large numbers:

$$\mu\left\{\mathbf{a} \in \{0, 1\}^{\omega}; \lim_{n \rightarrow \infty} (n^{-1}S_n(\mathbf{a})) = \frac{1}{2}\right\} = 1, \quad (3.7)$$

which says that the set defined within the braces in (3.7) is an element of \mathcal{L} (in fact even \mathcal{B}) and its Lebesgue measure is 1; hence in almost every infinite 0–1 sequence half the elements are zero and half one. This is a direct extension of the counting intuition expressed by the weak LLN (3.3). Indeed, the strong LLN (3.7) is a logical consequence of the weak law (3.3) in conjunction with σ -additivity. This means that the finite (3.3) and infinite (3.7) express the same idea, and σ -additivity is a way to translate the cumbersome (3.3) to the compact (3.7). Borel, the author of the strong LLN, actually preferred (3.3), in line with his intuitionistic views. He thought that (3.7) added nothing except for the illusion that infinite sets of infinite sequences made sense.

Similar observations can be made with respect to other limit laws that have familiar infinite formulations in \mathcal{L} , but also parallel formulations in \mathcal{F} which together with σ -additivity imply the infinite laws. An important example is the Law of Iterated Logarithm (LIL), a stronger and more subtle law than (3.7), which implies, among other things, that for almost every $\mathbf{a} \in \{0, 1\}^{\omega}$ the sign of n^{-1}

⁵The construction of \mathcal{B} is achieved by transfinite induction over the two operations, countable union and then countable intersection, all the way to the first uncountable ordinal.

⁶Further extensions of the Lebesgue measure are possible. The validity of the strong version of the axiom of choice entails the existence of non-measurable sets, that is, $C \subset \{0, 1\}^{\omega}$ such that $C \notin \mathcal{L}$. We can add some of those to \mathcal{L} and extend the measure to them [14]. With this the measure is no longer regular (see below). Moreover, there are models of set theory, with weaker principles of choice, in which every subset of $\{0, 1\}^{\omega}$ is Lebesgue measurable [15].

$S_n(\mathbf{a}) - 0.5$ oscillates infinitely often as $n \rightarrow \infty$. Sometimes the infinite law is more easily discovered than its finite parallel which may be even hard to formulate. In any case one can prove the regularity of μ , that every set in \mathcal{L} can be approximated by a set in \mathcal{F} to an arbitrary degree.

Theorem 1. *Let $E \in \mathcal{L}$ be any Lebesgue measurable set and let $\varepsilon > 0$; then there is $F_\varepsilon \in \mathcal{F}$ such that $\mu[(E \setminus F_\varepsilon) \cup (F_\varepsilon \setminus E)] < \varepsilon$.*

The proof is in Appendix 1 (note that the theorem becomes trivial when $\mu(E) = 0$ or $\mu(E) = 1$). Therefore, the expansion of the measure from the finite to the infinite domain conserves the meaning of the counting terms. We can, in principle, replace any set in $E \in \mathcal{L}$ by a finite set $F_\varepsilon \in \mathcal{F}$ which is arbitrarily close to E . If direct counting shows that F_ε comprises 0.75 of the cases, then so does E up to a small error.⁷ Moreover, the Lebesgue algebra \mathcal{L} is the maximal extension of \mathcal{F} for which theorem (1) is valid (see footnote 6). This seems to me to be a compelling argument for why \mathcal{L} is the correct extension of \mathcal{F} , and why the Lebesgue measure μ on \mathcal{L} is the correct extension of the combinatorial counting measure to infinity. It is also a compelling argument for why the notions of σ -algebra and σ -additivity are the appropriate tools in extending the combinatorial measure to infinity.

Let us come back to the issue of the Lebesgue measure and probability. As noted before only in rare cases do teachers make a point of distinguishing the meanings of weak LLN as a combinatorial and as a probabilistic statement. As for the strong LLN and other similar theorems, teachers and textbooks alike never make the distinction, and invariably interpret the Lebesgue measure in this context as probabilistic. There is no intrinsic reason for this, the application of σ -additivity has no probabilistic qualities. The reason is more sociological: For the pure mathematician there is no difference between the uniform probability distribution and the combinatorial measure, since their formal properties are one and the same. At a certain point in time mathematicians started to use the probabilistic language exclusively, and fellow scientists, physicists in particular, followed in their footsteps. But there is all the difference in the world between the mathematicians who are using the measure probabilistically, as a mere formality, and the physicists who are committing themselves to an application of probability as part of a theory of reality.

This has not always been the case, even for mathematicians! For example, in the struggle to obtain the correct estimation of frequency oscillations (LIL- the law of iterated logarithm), bounds were suggested by Hardy and Littlewood in 1914. They viewed the problem as number theoretic, concerning the binary development of real numbers between zero and one, and related to Diophantine approximation. Even in his final formulation of LIL from 1923 (for the uniform case) Khinchine was using the number-theoretic language, and only a year later switched to probability [16].

Extending the notion of vast majority from the finite to the infinite realm results in typical cases. None of these concepts is intrinsically probabilistic. I believe that

⁷ Actually $F_\varepsilon = A \times \{0, 1\} \times \{0, 1\} \times \dots \subseteq \{0, 1\}^\omega$ with $A \subseteq \{0, 1\}^n$ for some integer n , and we are counting the elements of A .

this is an important step towards removing the host of problems associated with probability distributions over initial conditions. As an example consider a recent application that does not even involve dynamics. Let a quantum system (“the universe”) be associated with a finite dimensional Hilbert space \mathcal{H} , with a large dimension D . Now, consider a small subsystem of dimension $d \ll D$ that corresponds to a subspace \mathcal{H}_1 . We can write $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$, where \mathcal{H}_2 is the Hilbert space of the environment with a large dimension $d^{-1}D$. The set of pure states in \mathcal{H} is the unit sphere of \mathcal{H} ; let μ be the normalized Lebesgue measure on it. Each pure state induces a mixed relative state on the small subsystem. The following recent result was proved independently in [17, 18]: *Almost all pure states in \mathcal{H} induce on \mathcal{H}_1 a relative state which is very close, in the trace norm, to the maximally mixed state on \mathcal{H}_1 , that is, $d^{-1}I_d$ with I_d the unit operator on \mathcal{H}_1 .*

One possible reading is that with probability one the state of the large system induces the near uniform state on the subsystem.⁸ A natural question is, “What does probability mean in this context?” Assume the large system is a model of the universe; it began in one pure state, and after time t it is again in one particular pure state. This state has been deterministically developed from the initial condition by the unitary time transformation. So the question is, “What do we mean by saying that the initial condition of the universe was picked from a uniform rather than some other probability distribution?” The only sensible answer is that this statement represents the epistemic probability of an agent who has no knowledge at all about the initial condition. However, this agent cannot be a physicist, who usually knows something about the present and earlier (macroscopic) states of the universe.

In the typicality approach, by contrast, the result simply means that the vast majority of pure states of the big system have the property in question, a combinatorial claim. This claim gives rise to a weak, but still informative conditional statement: If the universe began from a typical state then equilibrium should be a widespread phenomenon. A simple assumption (typicality) explains a large set of observations.

3.4 Dynamics

Our aim is to discuss the dynamical conditions that are the infinite parallels of the constraints **1, 2, 3** we have imposed on the movie. To fix notations let Γ denote the energy hypersurface of the closed system under consideration. If $x_0 \in \Gamma$ is a point, it can be considered as a possible initial condition, let $x_0(t)$ denote the trajectory starting from this point in Γ . Alternatively, if t is fixed $x_0(t)$ is the point to which x_0 travels after time t . The Lebesgue measure on Γ will be denoted by μ , and we assume it is normalized (we ignore the difficulties arising from a non compact Γ , which are settled by known techniques). The σ -algebra of the Lebesgue measurable

⁸ See [17]. In a later important dynamical extension of this result the authors adopt the typicality point of view [19].

sets will again be denoted by \mathcal{L} . If $E \in \mathcal{L}$, define E_t to be the time translation of E , that is, $E_t = \{x_0(t); x_0 \in E, \}$ for $0 \leq t < \infty$.

The second assumption **2** corresponds to the determinism inherent in classical mechanics and already reflected in the notation. The classical dynamical rule closest to assumption **1** is the conservation of energy. In the case of an ideal gas the velocities of the individual particles are varied but the average (square of the) particle's speeds remains constant (by analogy, the pace of the movie is constant). Energy conservation, that is, the Hamiltonian character of the system, also guarantees that the dynamics is measure preserving: $\mu(E) = \mu(E_t)$. In the movie case measure preservation is trivial.

Condition **3** corresponds to ergodicity. Historically, a major difficulty was associated with the formulation of this condition, Boltzmann mistakenly thought that a path can fill the whole energy hypersurface in phase space, so that every state will be visited. However, this requirement contradicts basic topological facts.⁹ It took a long struggle until the modern version of the ergodic condition was formulated, and the ergodic theorems subsequently proved [16]. Instead of referring to individual points visited by the path, the condition takes (measurable) set of points, and puts a constraint on the way the set fills up the space. Let E be a measurable subset of the energy hypersurface in phase space. Then E is *invariant* if for some $t > 0$ we have $E_t \subseteq E$. The system is ergodic if all invariant sets have measure zero or one.

In the finite case the dynamical rules provide an explanation why, in the long run, the movie is extremely boring and looks almost always gray. They also explain why, in the long run, the frequency of the pictures that have more black than white pixels is (a little less than) 0.5. This corresponds, in the infinite case, to the identity of the long run averages and the phase space averages of thermodynamic observables, a highly non-trivial fact which is the content of the ergodic theorems. In both cases the long run may be very long, in the infinite case there is no a priori bound on its length. This is the explanation why the system is at maximal entropy most of the time, or why about half the time the pressure in the left half of the container is less (even very slightly so) than in the right half.

However, there seems to be a difference between the finite and infinite case here. Given a thermodynamic observable, only typical initial conditions result in the identity of its phase space and long time averages. This may seem like a major difference from the finite case in which all initial conditions behave properly. However, a small amendment to the movie story can lead us to the conclusion that the movie satisfies condition **3** only for a vast majority of initial conditions, not all. To see this imagine that the set of pictures is divided into two disjoint subsets, one very small containing 20 pictures and the other containing the rest. When a movie begins with a picture in the small subset it goes through a small loop,

⁹Dimension is a topological invariant, as proved by Brouwer in 1911. Partial results concerning the non-existence of a homeomorphism between the real line and higher dimensional real spaces existed in Boltzmann's time. For example, Lüroth in 1878.

visiting all 20 pictures and starts again. Similarly for an initial condition in the second set, but then it covers all the pictures except 20. In both cases determinism is satisfied. We can say that for the vast majority of initial conditions the time and space averages of “thermodynamic observables”, functions $f: \{0, 1\}^{10^6} \rightarrow \mathbf{R}$, are (very nearly) the same.

It must be emphasized that the sense of explanation obtained in this manner is significant but limited. As a result of the unbounded nature of the long run, and in the absence of more information, there is no way we can combine the dynamical rules with the combinatorial facts to yield a definite prediction, for example, about what will take place 2 days from now. The kind of explanation we do have is weaker, and has the conditional form: “If the initial condition is typical, then. . . ” The assumption of typicality explains why the (calculated) space averages of observables are the same as the measured long time averages (which stabilize quickly in practice). Thus, assuming we are on a typical trajectory, one of a vast majority, explains much of what we actually see.

So far the explanation relies on the dynamical rules and the observations derived from the combinatorial nature of the Lebesgue measure. One may object to the latter point on the ground that the measure here does not seem to be “the same” as the measure on the set of infinite 0–1 sequences, being a Lebesgue measure on a Euclidean manifold of high dimension. This objection can be answered on two levels, the first is purely formal. As indicated before, all Lebesgue spaces which are defined on compact subsets of real or complex Euclidean spaces are isomorphic (after normalization of the measure) to the interval $[0,1]$ with the Lebesgue measure on it. Therefore, they are also isomorphic to the space of all 0–1 sequences, and every measurable set $E \subseteq \Gamma$ corresponds to a measurable set $\widehat{E} \subseteq \{0, 1\}^\omega$, with the same measure, and \widehat{E} can be approximated by a finite set $F \in \mathcal{F}$ as indicated in theorem 1.

On a deeper level there often exists a connection between ergodic systems and the sequence space when we apply a mapping of the ergodic system, including its dynamics, to the set two sided infinite 0–1 sequences [12, page 274]. This space, denoted by $\{0, 1\}^{\mathbf{Z}}$, is equipped with the (uniform) Lebesgue measure, and its elements can be written as $\mathbf{a} = (\dots, a_{-2}, a_{-1}, a_0, a_1, a_2, \dots)$, with $a_i \in \{0, 1\}$, $i = 0, \pm 1, \pm 2, \dots$. To perform the mapping between the thermodynamic system and this space one has to replace the continuous time variable by a discrete parameter. It turns out that many important ergodic systems, including the few physically realistic systems for which ergodicity was actually proved, are isomorphic as dynamical systems to the Bernoulli shift on $\{0, 1\}^{\mathbf{Z}}$, defined by¹⁰ $(\mathbf{S}\mathbf{a})_i = a_{i-1}$. These results were proved in a sequence of papers, mainly by Ornstein and his collaborators [12]. Ergodic systems with this property include the standard model of the ideal gas (hard-sphere molecules in a rectangular box),

¹⁰The reason why the double sided sequence space is used is to make the Bernoulli shift well defined and invertible.

Brownian motion in a rectangular region with reflecting boundary, geodesic flows in hyperbolic and many other spaces.

The connection with the combinatorial character of the measure is even more transparent in this case. For example, the Ergodic theorem for $\{0, 1\}^{\mathbb{Z}}$ with the shift entails the strong LLN. To see this let $A \subseteq \{0, 1\}^{\mathbb{Z}}$ be a measurable set then the ergodic theorem for the Bernoulli shift states,

$$\mu \left\{ \mathbf{a} \in \{0, 1\}^{\mathbb{Z}}; \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \chi_A(S^i(\mathbf{a})) = \mu(A) \right\} = 1. \quad (3.8)$$

Here χ_A is the indicator function of A , so that $\chi_A(\mathbf{a}) = 1$ if $\mathbf{a} \in A$, and $\chi_A(\mathbf{a}) = 0$ otherwise. Now take $A = \{\mathbf{a} \in \{0, 1\}^{\mathbb{Z}}; a_0 = 1\}$, then $\mu(A) = 0.5$ and $\sum_{i=1}^n \chi_A(S^i(\mathbf{a})) = \sum_{i=1}^n a_i$, and we obtain the strong LLN as a special case.

Probabilistic considerations enter when definite predictions are made, beyond the weaker long term explanations that are possible on the basis of ergodicity. Given the deterministic nature of the system we shall take probability in this context to be epistemic, although this may be disputed [7, 20]. The assignments of probabilities are based on knowledge about the system that may go beyond the simple rules we have considered. Some-times, in the absence of any knowledge about the initial condition and the dynamics beyond ergodicity, the uniform Lebesgue measure can serve as the degree of knowledge regarding the system. Often more knowledge is available, which can be theoretical, but frequently concerns the initial condition and is based on experience. For example, we may know something about the rate with which the dynamics is moving to mix the molecules. Usually the rate cannot be derived directly on the basis of the interactions between the particles. Higher theories such as fluid dynamics may be involved, together with experimental data. If a gas is prepared in a container with a divider, and the pressure on the left hand side much higher than the pressure on the right, then upon removing the divider the pressures will equalize very swiftly. By contrast, when we drop ink into water we know that it will take much longer to mix uniformly with the medium. Therefore, if we were to bet whether the pressures on both sides will equalize 20 s from now, the answer will be yes with probability close to 1, but the probability that the ink will be well mixed within 20 s is near zero. This does not follow from ergodicity which just explains why the system will eventually arrive at equilibrium and stay there most of the time.

We also know that in all recorded human history the reverse of these processes has never been reported. Consequently, the probability assigned to a spontaneous large pressure differences occurring within the next week (or month, or year. . .) is zero or very nearly so. This observation too cannot be derived logically from the dynamical and combinatorial rules. Given ergodicity, almost all initial conditions will take the system arbitrarily near every possible state. How do we know that the creation of a spontaneous large pressure difference is not around the corner?

We do know from combinatorial considerations that non equilibrium states are very rare, but this condition is insufficient to derive the probabilistic conclusion,

because we do not know what the trajectory is, and have no clue about the way rare states are distributed on it. The movie analog is a photograph of the Empire State Building appearing suddenly in the midst of gray pictures. This photograph must appear sometime, but in the absence of detailed knowledge of the dynamics one cannot tell when. However, after sitting 10^{10} years and watching gray pictures one may assign the sudden pop up of the Empire State Building in the next week a very small probability. This would not be the case after a long stretch of pictures of buildings. By analogy, we assign zero probability to the creation next week of a spontaneous large pressure differential because this has never happened, and not just because we know abstractly that this is an atypical event.¹¹

3.5 Troubles with Typicality

The problem is that typicality is too restrictive a notion, and the reasons are twofold, physical and logical. Physically, there are good reasons to deal with measurable sets of intermediate size. For example, the set of micro states for which the pressure in the left half of the container is equal or less than the pressure in the right comprise 0.5 of all the states. Logically, we shall see that the concept of typicality lacks closure. For example, even after typical points have been “fixed” one cannot use this stipulation to define typical pairs of points, that is, *a pair of typical points is not necessarily a typical pair of points*. To define the latter, one has to go back to the Lebesgue measure on the set of pairs (which is defined in terms of the Lebesgue measure on the set of singletons) and redefine typicality for pairs.

As for the physical restriction, one important case is that of smooth classical Hamiltonian systems which are not ergodic, but only measure preserving. By Birkhoff’s theorem convergence of the time average of a thermodynamic observable for typical initial conditions is guaranteed, but the result is not identical to the space average. In this case the phase space is partitioned into invariant sets of positive measure, such that the restriction of the dynamics to each element in the partition is ergodic (after a suitable renormalization). By KAM’s theorem many Hamiltonian systems are not ergodic, although the partition is often composed of one large invariant set and other much smaller elements. (For such systems the notion of ε ergodicity has been introduced [22]). Even in this case one has to say something about sets of initial conditions with measure smaller than 1, which cannot even be formulated without the full Lebesgue measure.

The logical point is that exchanging the full Lebesgue measure for the weaker notion of typicality does not even accomplish the task of explaining the long run statistical regularities. In order to provide such an explanation one has to introduce an infinite sequence of logically independent concepts of typicality, none of which

¹¹ A similar point about the role of induction in statistical mechanics is made in [21].

are definable in terms of the former. Consider Galton's board, which serves as a central example in the papers by Dürr [3] and Maudlin [8]. The first notion introduced is that of a typical initial condition, which explains, e.g., the stability of relative frequencies of going left and going right. Next, we must introduce a new notion of typical pairs of initial conditions to explain the stability of the frequency of the correlated sequence obtained from two runs of the board, then we have to introduce a new notion of typical triples to explain the stability of triple correlated sequences obtained from three runs, and so on. Each one of these notions is logically independent of the former notions, that is, none of them can be defined on the basis of the previous concepts of typicality. In each case one has to reintroduce the fully fledged Lebesgue measure (respectively, on the interval of initial conditions, the Cartesian product of the interval by itself, the three-fold Cartesian product, and so on), and only then, in each case separately, throw away the ladder as it were, and introduce the new notion of typicality in the manner described by Maudlin for the singleton case.

One consequence of this state of affairs is that being typical is not an intrinsic property of a point even for a single dynamical system, but is a property induced by its relations to other points. Moving to the system comprising the whole universe (which after all has only one initial state) does not solve the problem. In this case it also arises in the context of the typicality of idealized sequences of empirical observations, the correlations or independence of two such sequences, and of triples, etc. Even if we observe only one (ideally infinite) typical sequence, the problem arises with respect to its subsequences and their relations.

To see this consider a pair of $\mathbf{a}, \mathbf{b} \in \{0, 1\}^\omega$ and denote $\mathbf{a} \cdot \mathbf{b} = (a_1 b_1, a_2 b_2, a_2 b_2, \dots)$. We know that typically $\mathbf{a} \cdot \mathbf{b}$ is a sequence whose averages satisfy $\frac{1}{n} \times \sum_{i=1}^n a_i b_i \rightarrow 0.25$. But does this fact follow if we assume that \mathbf{a} and \mathbf{b} are typical? The negative answer follows from

Theorem 2. *Let $A \subset \{0, 1\}^\omega$ be any measurable set with $\mu(A) > \frac{1}{2}$; then there are $\mathbf{a}, \mathbf{b} \in A$ such that $\mathbf{a} \cdot \mathbf{b}$ has a divergent sequence of averages.*

The proof is in Appendix 2. This means that no matter what the set of typical sequences is, there will always be pairs of typical sequences whose correlation is not even defined. One might object on the ground that the set of such bad pairs has measure zero, and the set of typical pairs has measure one. However, this refers to the measure *on the Lebesgue space of pairs*. The set of typical pairs does not have the form $A \times A$ with $A \in \mathcal{L}$, and $\mu(A) = 1$. By theorem 2 any set of the form $A \times A$ contained in the set of typical pairs has at most measure $\mu(A) \leq 0.5$. Therefore, to be able to speak about typical pairs one has to construct first the Lebesgue measure on the set of pairs $\{0, 1\}^\omega \times \{0, 1\}^\omega$, or alternatively $[0, 1] \times [0, 1]$, and only then define typicality for pairs. One cannot do it by relying on the already established set of typical points. This observation can be extended to triple, quadruple correlations, and so forth. In the case of triples the equivalent theorem applies when $\mu(A) > \frac{1}{3}$, and so on, for k -tuples when $\mu(A) > \frac{1}{k}$. In all these cases the notion of typicality cannot be derived from the lower dimensional ones.

As noted this also means the being typical is not an intrinsic property of an initial condition, not even for a single fixed system, but depends on the relation between the point and other possible initial conditions. The way suggested here to avoid this difficulty is to use the fully fledged Lebesgue measure, in its combinatorial interpretation. In this case subsets of measure one are just special cases. I think all the advantages of the concept of typicality that were pointed out in the literature are preserved, but the difficulties are avoided.

Acknowledgement *I would like to thank Meir Hemmo and Orly Shenker for their valuable advice. This research is supported by the Israel Science Foundation, grant number 744/07.*

Appendix 1: Proof of Theorem 1

Theorem 1. *Let $E \in \mathcal{L}$ be any Lebesgue measurable set and let $\varepsilon > 0$; then there is $F_\varepsilon \in \mathcal{F}$ such that $\mu[(E \setminus F_\varepsilon) \cup (F_\varepsilon \setminus E)] < \varepsilon$.*

Proof. Consider first a Lebesgue measurable subset $E \subseteq [0, 1]$. By the regularity of the Lebesgue measure (see [23], page 230) given any $\varepsilon > 0$ there is an open set U , with $E \subseteq U$ and $\mu(U \setminus E) < \frac{\varepsilon}{2}$. The family of open intervals with dyadic endpoints forms a basis for the usual topology on $[0, 1]$ (recall that a dyadic number is a rational whose denominator is a power of 2). Thus, we can represent U as a disjoint countable union $U = \cup_{j=1}^{\infty} (c_j, d_j)$, where c_j and d_j are dyadic, and $\mu(U) = \sum_{j=1}^{\infty} (d_j - c_j)$. By choosing a sufficiently large natural number N we can make sure that $U' = \cup_{j=1}^N (c_j, d_j) \subseteq U$ satisfies $\mu(U') > \mu(U) - \frac{\varepsilon}{2}$. Now define U'' to be the set obtained from U' by adding the endpoints of each interval: $U'' = \cup_{j=1}^N [c_j, d_j]$. Since we have added just finitely many points the measure of U'' is the same as that of U' , and therefore, $\mu[(E \setminus U'') \cup (U'' \setminus E)] < \varepsilon$.

Now apply the map $\sum_{j=1}^{\infty} a_j 2^{-j} \rightarrow (a_1, a_2, a_3, \dots)$ which takes real numbers in $[0, 1]$ to their sequence of binary coefficients in $\{0, 1\}^{\omega}$. Dyadic rationals have two binary developments, one ending with an infinite sequence of zeroes, and the other ending with an infinite sequence of ones. Adopt the convention that in case of a dyadic rational d , the map takes d to its two binary sequences. Since the set of dyadic numbers has measure zero the map is measure preserving. The set E is then mapped to a subset of $\{0, 1\}^{\omega}$ which we shall also denote by E . The set U'' is mapped to a *finite* subset of $\{0, 1\}^{\omega}$ which we will denote by $F_\varepsilon \in \mathcal{F}$. The reason is that every closed interval with dyadic endpoints is mapped to a finite set, for example, $[\frac{1}{4}, \frac{5}{8}] \rightarrow \{(0, 1, 0), (0, 1, 1), (1, 0, 0)\} \times \{0, 1\} \times \{0, 1\} \times \dots \subseteq \{0, 1\}^{\omega}$, and U'' is a finite union of such intervals. This completes the proof.

Appendix 2: Proof of Theorem 2

This theorem and proof appeared first in [24] as part of a criticism of the frequency interpretation of probability.

Theorem 2. *Let $A \subset \{0, 1\}^\omega$ be any measurable set with $\mu(A) > \frac{1}{2}$, then there are $\mathbf{a}, \mathbf{b} \in A$ such that $\mathbf{a} \oplus \mathbf{b}$ has a divergent sequence of averages.*

Proof. Denote by $\mathbf{a} \oplus \mathbf{b}$ the XOR of the elements \mathbf{a} and \mathbf{b} , in other words $(\mathbf{a} \oplus \mathbf{b})_i = a_i + b_i \pmod{2}$. We first show that $\mu(A) > \frac{1}{2}$ implies that $A \oplus A = \{\mathbf{a} \oplus \mathbf{b}; \mathbf{a}, \mathbf{b} \in A\} = \{0, 1\}^\omega$. Indeed if $\mathbf{c} \notin A \oplus A$, then $(\mathbf{c} \oplus A) \cap A = \emptyset$, where $\mathbf{c} \oplus A = \{\mathbf{c} \oplus \mathbf{a}; \mathbf{a} \in A\}$. Otherwise, if $\mathbf{d} \in (\mathbf{c} \oplus A) \cap A$ then $\mathbf{d} \in A$ and $\mathbf{d} = (\mathbf{c} \oplus \mathbf{a})$ for some $\mathbf{a} \in A$. Hence $\mathbf{c} = (\mathbf{d} \oplus \mathbf{a}) \in A \oplus A$, contradiction. Therefore, $(\mathbf{c} \oplus A) \cap A = \emptyset$, but this also leads to a contradiction since $\mu(\mathbf{c} \oplus A) = \mu(A) > \frac{1}{2}$, hence $A \oplus A = \{0, 1\}^\omega$.

We can assume without loss of generality that all elements of A have a convergent sequence of averages. This is the case because the set of elements of $\{0, 1\}^\omega$ whose averages diverge has measure zero. Let $\mathbf{c} \in \{0, 1\}^\omega$ be some sequence with a divergent sequence of averages. Then by the above argument there are $\mathbf{a}, \mathbf{b} \in A$ such that $\mathbf{c} = (\mathbf{a} \oplus \mathbf{b})$, that is $c_i = a_i + b_i \pmod{2} = a_i + b_i - 2a_i b_i$ and therefore

$$\frac{1}{n} \sum_{i=1}^n c_i = \frac{1}{n} \sum_{i=1}^n a_i + \frac{1}{n} \sum_{i=1}^n b_i - \frac{2}{n} \sum_{i=1}^n a_i b_i.$$

The sequence on the left diverges, and the first two sequences on the right converge. Hence, $n^{-1} \sum_{i=1}^n a_i b_i$ diverges. This completes the proof.

References

1. Lebowitz, J.L.: Boltzmann's entropy and time's arrow. *Phys. Today* **46**, 32–38 (1993)
2. Lebowitz, J.L.: Macroscopic laws, microscopic dynamics, time's arrow and Boltzmann's entropy. *Physica A* **194**, 1–27 (1993)
3. Dürr, D.: Über den Zufall in der Physik. <http://www.mathematik.uni-muenchen.de/~duerr/Zufall/zufall.html> (1998)
4. Goldstein, S.: Boltzmann's approach to statistical mechanics. In: Jean, B., Dürr, D., Galavotti, M.C., Ghirardi, G.C., Petruccione, F., Zangh, N. (eds.) *Chance in Physics: Foundations and Perspectives*, pp. 39–54. Springer, Berlin and New York (2001)
5. Goldstein, S., Lebowitz, J.L.: On the (Boltzmann) entropy of non-equilibrium systems. *Physica D* **193**, 53–66 (2004)
6. Lavis, D.: Boltzmann and Gibbs: an attempted reconciliation. *Stud. Hist. Philos. Mod. Phys.* **36**, 245–273 (2005)
7. Maudlin, T.: What could be objective about probabilities? *Stud. Hist. Philos. Mod. Phys.* **38**, 275–291 (2007)
8. Sheldon, G., Lebowitz, J.L.: Mastrodonato Christian, Tumulka Roderich, Zanghi Nino. Normal Typicality and von Neumann's Quantum Ergodic Theorem. *Proc. R. Soc. A* **466**(2123), 3203–3224 (2009)

9. Frigg, R.: Typicality and the approach to equilibrium in Boltzmannian statistical mechanics. In: Ernst, G., Huttemann, A. (eds.) *Time Chance and Reduction. Philosophical Aspects of Statistical Mechanics*, pp. 92–118. Cambridge University Press, Cambridge (2010)
10. Uffink, J.: Compendium of the foundations of classical statistical physics. In: Butterfield, J., Earman, J. (eds.) *Handbook for Philosophy of Physics*, pp. 923–1047. North Holland, Amsterdam (2006)
11. Buzaglo, M.: *The Logic of Concept Expansion*. Cambridge University Press, Cambridge (2002)
12. Patersen, K.: *Ergodic Theory*. Cambridge University Press, Cambridge (1983)
13. Chow, Y.S., Teicher, H.: *Probability Theory Independence, Interchangeability, Martingales*. Springer, Berlin and New York (1978)
14. Hewitt, E., Ross, K.A.: *Abstract Harmonic Analysis*. Springer, Berlin and New York (1994)
15. Solovay, R.M.: A model of set-theory in which every set of reals is Lebesgue measurable. *Ann. Math.* **92**, 1 (1970)
16. von Plato, J.: *Creating Modern Probability*. Cambridge University Press, Cambridge (1994)
17. Sandu, P., Short, A.J.: Winter Andreas. The foundations of statistical mechanics from entanglement: individual states vs. averages. <http://arxiv.org/abs/quantph/0511225> (2005)
18. Sheldon, G., Lebowitz, J.L., Tumulka, R., Zanghi, N.: Canonical typicality. *Phys. Rev. Lett.* **96**, 050403 (2006)
19. Noah, L., Sandu, P., Short, A.J.: Winter Andreas. Quantum mechanical evolution towards thermal equilibrium. <http://arxiv.org/abs/0812.2385> (2008)
20. Loewer, B.: Determinism and chance. *Stud. Hist. Philos. Mod. Phys.* **32**, 609–629 (2001)
21. Hemmo, M., Orly, S.: Measures over initial conditions, in Y. Ben-Menahem and M. Hemmo (eds.), *Probability in Physics*, pp. 87–98. The Frontiers Collection, Springer-Verlag Berlin Heidelberg (2011)
22. Vranas, P.B.M.: Epsilon-ergodicity and the success of equilibrium statistical mechanics. *Philos. Sci.* **65**, 688–708 (1998)
23. Rudin, W.: *Principles of Mathematical Analysis*. McGraw Hill, New York (1964)
24. Bub, J., Pitowsky, I.: Critical notice on Popper's postscript to the logic of scientific discovery. *Can. J. Philos.* **15**, 539–552 (1986)