Received: 6 July 2017 | Accepted: 7 August 2017

DOI: 10.1111/phc3.12465

WILEY

ARTICLE

Foundation of statistical mechanics: Mechanics by itself

Orly Shenker

The Hebrew University of Jerusalem

Correspondence
Orly Shenker, Edelstein Centre for History and Philosophy of Science, The Hebrew University of Jerusalem, Edmund J. Safra Campus, Jerusalem 91904, Israel.
Email: orly.shenker@mail.huji.ac.il

Abstract

Statistical mechanics is a strange theory. Its aims are debated, its methods are contested, its main claims have never been fully proven, and their very truth is challenged, yet at the same time, it enjoys huge empirical success and gives us the feeling that we understand important phenomena. What is this weird theory, exactly? Statistical mechanics is the name of the ongoing attempt to apply mechanics (classical, as discussed in this paper, or quantum), together with some auxiliary hypotheses, to explain and predict certain phenomena, above all those described by thermodynamics. This paper shows what parts of this objective can be achieved with mechanics by itself. It thus clarifies what roles remain for the auxiliary assumptions that are needed to achieve the rest of the desiderata. Those auxiliary hypotheses are described in another paper in this journal, Foundations of statistical mechanics: The auxiliary hypotheses.

1 | INTRODUCTION

Statistical mechanics is a strange theory. Its aims are debated, its methods are contested, its main claims have never been fully proven, and their very truth is challenged, yet at the same time, it enjoys huge empirical success and gives us the feeling that we understand important phenomena. What is this weird theory, exactly? Statistical mechanics is not on a par with (so-called fundamental) theories like classical or quantum mechanics. Rather, it is the name of the ongoing attempt to *apply* such theories, together with some other premises (described below), to explain and predict certain phenomena, above all those described by thermodynamics.[1] Einstein (1919) portrayed this nature of statistical mechanics when he distinguished between *theories of principle* and *constructive theories*.[2] The former, he said, describe "empirically observed general properties of phenomena," and he took thermodynamics to be their paradigmatic example. The latter, he added, explain these phenomena "out of some relatively simple propositions," its paradigm being statistical mechanics. However, there was never such a clear distinction between the *explanandum* and the *explanans* in this field. While it has always been clear that the explanandum has to include the thermodynamic phenomena and the explanans has to include the fundamental theories of physics, the creators of statistical mechanics understood early on that the laws of thermodynamics *cannot* be deduced from the fundamental theories of physics

alone. After all, it is a point of *logic* that time-asymmetric conclusions, such as the second law of thermodynamics, cannot be validly derived from time-symmetric assumptions, such as the laws of the fundamental theories.[3] Therefore, the creation of statistical mechanics involved redefinitions of the explanandum as well as the explanans, in such a way that the *new* explanans—now consisting of mechanics together with some supplements—entails the *new* explanandum—now no longer the original thermodynamic laws.

The main revision to the explanandum was abandoning the idea that the laws of thermodynamics are universal and instead accepting that the thermodynamic regularities are (merely) *highly likely* or *very typical*.[4] The earliest successful attempts to recover the *synchronous* thermodynamic regularities (namely, the relations between thermodynamic quantities *at a time*) from mechanics were already based on probabilistic assumptions. Famously, when Maxwell investigated the paradigmatic case of an ideal gas in equilibrium,[5] he made *statistical* assumptions about the distribution of velocities among the gas particles, which he deemed "precarious" (see Uffink, 2007). Maxwell's statistical assumptions, which later became part of the famous Maxwell–Boltzmann energy distribution, entail that certain predictions concerning thermodynamic magnitudes are, at best, only highly probable. Taking these predictions to be successful meant accepting (implicitly) that probabilistic statements are the new desiderata. It soon became clear that the *diachronous* second law of thermodynamics posed an even greater challenge, as it was quickly realized that it is provably impossible to derive the absolute and universal approach to equilibrium from mechanics (due to Loschmidt's reversal theorem, Zermelo and Poincaré's recurrence theorem, and Liouville's theorem). There seemed, however, to be no a priori obstacle to proving that the approach to equilibrium is highly probable, and the conclusion was endorsing a new explanandum, which claims that the thermodynamic regularities are (merely) *highly likely* or *very typical*. The meaning of these probability and typicality statements are still under debate, and this means that the theory's explanandum is still disputed (a brief overview of this debate is in Shenker, 2017, Section 3).

What is the explanans that suits this explanandum? Since the fundamental theories of physics are a major part of the explanans, their character clearly affects the theory. Much of the study of the foundations of statistical mechanics is carried out under the assumption that classical mechanics is the fundamental theory. Although strictly speaking, according to contemporary physics, classical mechanics is false, it arguably preserves some of the salient explanatory and predictive aspects of the true fundamental physics, under the appropriate conditions.[6] With these caveats in mind, the subject of this article is the foundations of *classical* statistical mechanics (hereafter, for brevity: *statistical mechanics*). (Emch, 2007 is an overview of quantum statistical mechanics, while Shenker, 2018, is an overview of philosophical issues in the foundations of quantum statistical mechanics.)

This article will discuss the conceptual bridge between the explanandum and explanans. After setting the stage (in Section 2) by describing the basic ontology of classical mechanics, focusing on the notions of microstates and their aspects (often referred to in terms such as macrovariables, macroconditions, and macrostates), I describe (in Section 3) the way in which thermodynamic magnitudes are associated with mechanical ones, and then (in Section 4) explain how probabilistic statements can appear within classical mechanics. I conclude (in Section 5) by reflecting on what mechanics by itself can provide and where additional assumptions may be needed in the explanans in order to achieve the desired explanandum. Those auxiliary hypotheses are discussed in Shenker (2017).

There is excellent literature on the history and state of art of the foundations of statistical mechanics (e.g., Frigg, 2008; Uffink, 2007). Writing a new article on the subject (such as the present one) can only be justified if it reports some of the work that has been done in the field since the existing reviews were published or offers a different perspective on some major ideas in the field. In this article, I hope to do both.

2 | THE BASIC MECHANICAL ONTOLOGY

According to classical mechanics, at every moment, the universe is in some well-defined state called a *microstate*, consisting of the positions and velocities of all the particles.[7] Given the parameters and constraints (such as the

masses of the particles, and the total energy and volume available to them) and the microstate at some moment, the equations of motion determine the microstates at all other moments: The sequence of microstates that is formed in this way is often called a *trajectory* in the system's states space. The following terminological remark is of utmost importance for understanding statistical mechanics: There are various notions of *microscopic* in the literature; the term sometimes means *small*, and sometimes *part of a whole*.[8] But in statistical mechanics, it is customary to use the term "microstate" to denote the *complete* mechanical state of the system of interest, or of the world. It is complete in the sense that (given the parameters and constraints) the theory cannot say any more about the state of affairs in the world and does not need to have more and cannot settle for less,[9] in order to predict and retrodict other microstates. For instance, the momentary precise positions and velocities of the particles in the entire universe is a microstate, despite the fact that it is neither small nor part of anything. This meaning of *microstate* is pivotal for understanding the theory, as we shall see shortly.

"At first sight we might conclude … that, as the number of particles increases, so also must the complexity and intricacy of the properties of the mechanical system, and that no trace of regularity can be found in the behavior of a macroscopic body. This is not so, however: … when the number of particles is very large, new types of regularity appear." (Landau & Lifshitz, 1980, p. 1) Paramount of these new regularities are the thermodynamics ones. It was the greatest discovery of the creators of statistical mechanics, especially Maxwell (see Myrvold, 2011), Boltzmann (see Uffink, 2004), and Gibbs (1902) that only *partial information* about the microstate of a system is needed for providing a successful and informative account of thermodynamic phenomena.[10] Here is an example.

Consider a particular sample of a dilute gas (idealized as *ideal gas*) in a container, observed during some particular time interval from t_1 to t_2, and seen to stably fill the entire container with uniform pressure and temperature. We say that this sample of gas is in thermodynamic equilibrium. *Thermodynamic equilibrium* is a state in which the system has a certain set of thermodynamic properties, uniquely fixed by its constitution and the constraints on it, which do not change with time (see Brown & Uffink, 2001). We know from experience, described in *thermodynamics*, that the volume, pressure and temperature of an ideal gas in equilibrium are related according to the *ideal gas law* $PV = nRT$, where P is pressure, V is volume, and T is temperature (n is the number of moles and R is the gas constant; see Fermi, 1936). What do we know about this sample of gas in terms of *mechanics*? First of all, we know from mechanics that during this interval the gas particles evolve through a continuous sequence of *different* mechanical microstates. But we know a bit more. It has been discovered that the thermodynamic terms that appear in the ideal gas law have mechanical counterparts that satisfy the closely related functional relation $PV = NkT$ (where N is the number of particles and k is Boltzmann's constant), and that (as Maxwell and Boltzmann showed) in an ideal gas in equilibrium, the total mechanical energy is distributed among the particles according to the Maxwell–Boltzmann distribution (see overview of these discoveries in Sklar, 1993). And so we know that although our gas sample evolve through different microstates, they all *share an aspect*,[11] namely, in all of them, the energy is distributed among the particles according to the Maxwell–Boltzmann energy distribution.[12]

The notion of *aspect* has two features, whose interplay is the core of statistical mechanics: aspect as pertaining to an individual microstate and aspect as pertaining to a set of microstates. Portides (2017) uses a similar idea when he argues that a key element in the construction of scientific models is the extraction of relevant features of the system or process to be modeled, or selective attention to its features, rather than the substraction or change of features (as abstraction and idealization, respectively, are often understood). Ben Menahem (2001) argues that when focusing on different aspects of the same system, different regularities may appear; see Shenker (2014). As an example, let us focus our attention on one particular microstate out of the above sequence, which obtains at the point of time t_a between t_1 and t_2 (see Figure 1). We know that this particular microstate has the aspect of the Maxwell–Boltzmann energy distribution (call this aspect *MB*). When we say that at t_a, the system is in the Maxwell–Boltzmann energy distribution we are talking about an aspect, given by a partial description, of this *particular microstate*. However, this particular microstate also has other aspects, about which we know nothing (by hypothesis: e.g., the precise positions of the particles). Consequently, the partial information that we have concerning the t_a microstate is compatible with the system's being in any one of a continuous infinity of microstates that share this aspect but differ in their other

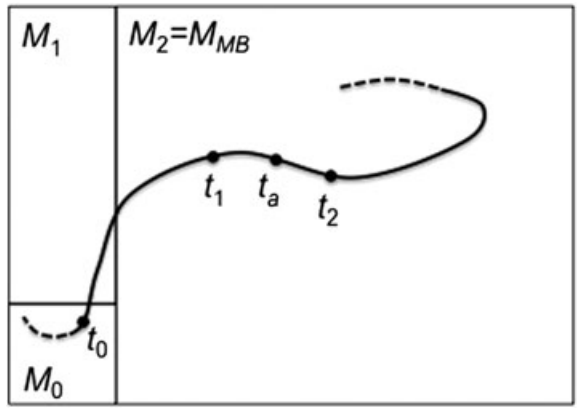

FIGURE 1 The mechanical account of thermodynamic regularities

aspects. All we know is that the actual microstate at t_a is a member of the *set* of microstates that have the aspect *MB* (call this set M_{MB}). One of the microstates in this set is *actual* (at t_a), and the rest are *counterfactual* (at t_a). But since the only thing we know about the actual microstates is that it has the aspect *MB*, we do not know which microstate out of the set M_{MB} the actual one is. (Some of the microstates of M_{MB} that are counterfactual at t_a will obtain in other moments, but many will *never* obtain for that system.) When we think of the partial description of our actual microstate at t_a (which pertains to the said aspect) as giving rise to a *set* of counterfactual microstates, this is often referred to in terms of "macro": The aspect itself is sometimes called a *macrovariable*, and the set of counterfactual microstates to which it gives rise is sometimes called a *macrostate*; sometimes, these terms are used interchangeably or other terms are used (see various examples in Ehrenfest & Ehrenfest, 1912; Sklar, 1993; Lebowitz, 1993; Albert, 2000; Goldstein & Lebowitz, 2004; Frigg, 2008).[13] (Gomori, Gyenis, & Hofer-Szabo, 2017, provide further formal analysis of how macrostates come about.) To avoid conceptual confusion, I will use the terms *aspect* (given by a partial description) and a *set* of microstates that share an aspect (instead of the terms macrovariables, macrostates, etc.).

As the system evolves and one microstate is replaced by the next according the equation of motion, the aspects of the microstates may change as well. In Figure 1, for example, when the system evolves from the microstate at t_1 to the microstate at t_2, the aspect *MB* remains unchanged, and all the microstates are in the set M_{MB}; but when the system evolves from t_0 to t_1, the aspects change as we move from M_0 via M_1 to M_{MB}. In these terms, the idea of the creators of statistical mechanics was that it turns out to be a *fact* that the microstates of big and complex systems, of the sort that is described by thermodynamics, have certain *aspects that exhibit regularities*. Let us see how these aspects are discovered and then describe their regularities in mechanical terms.

3 | THE MECHANICAL COUNTERPARTS OF THERMODYNAMIC MAGNITUDES

Every microstate has infinitely many aspects (given by partial descriptions), and (only) some of them are associated with thermodynamic magnitudes. The central idea of statistical mechanics is that the regularities that govern these mechanical aspects, according to the laws of mechanics, account for the thermodynamic regularities. Thus, to see how the laws of mechanics can explain the thermodynamic regularities, it is necessary to discover first which mechanical magnitudes are associated with the thermodynamic ones.

There are two competing ways of associating thermodynamic magnitudes with mechanical ones, offered by two theoretical frameworks, both called classical statistical mechanics. One of them follows ideas of Boltzmann and the other follows those of Gibbs (the two frameworks are described and compared in detail in for example Frigg, 2008

and Uffink, 2007). In the Boltzmannian approach, every thermodynamic magnitude is associated with some mechanical aspect shared by a set of microstates of the system during a given observation. When the system evolves, as long as its new microstate has the same aspect, the thermodynamic magnitude does not change; but when the aspect changes, the thermodynamic magnitude changes. In the Gibbsian approach, every thermodynamic magnitude is associated with a function of a set of microstates that have different aspects, weighted according to the (conjectured) relative frequencies of their occurrences during the observation. Therefore, the microstate may evolve in such a way that its aspect changes but the thermodynamic magnitude does not change. What justifies these two types of associations?

One might think that associating observable thermodynamic magnitude with mechanical aspect would be a matter of conceptual analysis, based on the idea that what we observe are effects of aspects of the microscopic furniture of the world; but historically, this was not the way these associations were discovered. In the Boltzmannian tradition, the association of certain aspects of microstates with certain thermodynamic magnitudes came to be endorsed following a generalization from the results in a special (albeit important) case. Early on Maxwell found that he could make successful predictions by associating the thermodynamic magnitude of equilibrium in an ideal gas with the mechanical magnitude of a certain distribution of velocities (and hence the kinetic energy) among the gas particles. Boltzmann later generalized this result to include other forms of energy, and the distribution of energy among the particles of an ideal gas in equilibrium came to be known as the Maxwell–Boltzmann energy distribution. Although Maxwell's rationale for his velocity distribution was unclear even to himself, it was endorsed because non-trivial predictions of thermodynamic magnitudes, derived on the basis of this association, were empirically confirmed. The empirical success of associating the Maxwell–Boltzmann energy distribution with thermodynamic equilibrium was later supplemented by Boltzmann's finding that this energy distribution satisfies certain extremum conditions (in his early *H*-theorem[14] as well as the later combinatorial argument) which suggest, conceptually, that this distribution is the unique equilibrium state to which systems evolve according to the second law of thermodynamics. A third reason for associating thermodynamic magnitudes with mechanical ones was the successful derivation of certain *functional relations* between various mechanical aspects that mirror the functional relations between thermodynamic magnitudes: associating magnitudes on the basis of such analogies proved successful as well; the case of the ideal gas law mentioned above is an example.[15] While the very idea that thermodynamic magnitudes should be identified with mechanical aspects is based on a general reductionist approach, the above discoveries gave it considerable support. An example that became famous in the philosophical literature for associating thermodynamic magnitudes with mechanical ones is the association of average kinetic energy with temperature: "Heat is the motion of molecules" (Kripke, 1980).[16]

Here is an important result of the Boltzmannian way of associating thermodynamic magnitudes with mechanical aspects: Since changing a mechanical aspect means changing the associated thermodynamic magnitude, and since it has been discovered that the change in the relevant mechanical aspects satisfies a probabilistic regularity (as discussed in the next section), in the Boltzmannian tradition, the laws of thermodynamics were replaced by probabilistic laws, which became the new explanandum in statistical mechanics. The Boltzmannian approach differs in this way from the Gibbsian one, which I will now discuss. (For more details about the Boltzmannian approach, see Frigg, 2008; Uffink, 2004, 2007; Werndl & Frigg, 2015a, b.)[17]

Gibbs (1902) discovered that certain *state-space functions* (e.g., averages of certain aspects in certain sets of microstates) satisfy functional relations that structurally parallel those that hold between certain thermodynamic magnitudes. Gibbs was explicitly cautious and avoided drawing strong ontological conclusions from these structural similarities, calling them "analogies." In practice, however, the Gibbsian state-space functions are often treated as giving rise to, or even identical with, the corresponding thermodynamic magnitudes. The conceptual justification often given for this is the following: Measurements take time, and during that time, the microstates change, in such a way that the relevant aspects occasionally change as well, with relative frequencies expressed by certain measures over the state space; and the measured thermodynamic magnitudes are the *total* effect exerted on the measuring devices during a measurement by those weighted aspects. An important result of this Gibbsian way of associating thermodynamic

magnitudes with mechanical aspects is this: While in Boltzmann's approach, the probabilistic regularity governing the microscopic evolution translates into a probabilistic regularity governing the thermodynamic magnitudes, in Gibbs's approach, this probabilistic regularity is absorbed, as it were, into the state-space averages, and the resulting laws, which are *about the averages*, remain absolute, not probabilistic. Those absolute laws are the Gibbsian explanandum. (See Frigg, 2008, and Wallace, 2015, for more details on the Gibbsian framework, and Werndl & Frigg, 2017 and Jaynes 1965 for the difference between its predictions and the Boltzmannian ones. See Ridderbos & Redhead 1998 and Callender, 1999, for its conceptual appraisal. Lavis, 2005, 2007, and Hemmo & Shenker, 2012, Chapter 11, attempt to explain the predictive success of Gibbs's approach despite its conceptual difficulties.)

Due to the differences between the details of the Boltzmannian and Gibbsian ways of associating the thermodynamic with the mechanical magnitudes, their predictions agree only approximately (Werndl & Frigg, 2017, discuss these predictive differences and their significance; see also Jaynes, 1965). However they are both applied to explain and predict the thermodynamic phenomena, and it is therefore important to see the connection between them (see Hemmo & Shenker, 2012, Chapter 11). Consider the measurement of the volume of a gas. During the measurement, at each moment, the measuring device interacts with *only some* of the gas particles. For example, photons coming from *certain* gas particles reach the device and begin a chain of reactions, by the end of which the device is in some pre-defined stable state. However, if we make some (reasonable, acceptable) statistical and dynamical assumptions about the observed gas (together with implicit assumptions concerning its interaction with the measuring device and the latter's internal mechanism), we can infer that whenever the gas passes through a certain kind of mechanical sequence of aspects, the device will register a suitable outcome. In Boltzmann's approach, the aspects in this sequence are all the same; in Gibbs's approach, they are different and appear with certain characteristic relative frequencies. On this understanding, both approaches provide *reductive* accounts of the identity statement between the thermodynamic magnitude of "volume" and the mechanical magnitude of "distribution of positions." Other, more complex identity statements, such as "Heat is the motion of molecules" (Kripke, 1980), are explained along the same reductive lines.[18]

4 | STATISTICAL MECHANICAL COUNTERPARTS OF THERMODYNAMIC REGULARITIES

Suppose we observe a system S_0, find that its microstate (at the time of observation t_0) has the mechanical aspect M_0, and want to predict its evolution. For example, we may want to predict whether its aspect at t_1 will be M_1 or M_2, given that its evolution is governed by a given Hamiltonian H. (This task sounds most natural within the Boltzmannian framework, but it is also applicable to the Gibbsian framework, using the connection between these frameworks described briefly above.) Our task is to do so on the basis of mechanics only. A convenient tool for thinking about this task is *Laplace's Demon*, an imaginary creature that has infinitely precise measurement capabilities and infinite calculation capabilities. Due to these unlimited capabilities Laplace's Demon personifies the theory of mechanics itself: It can do everything that mechanics can "do." While the capabilities of Laplace's Demon make it an idealization that does not even remotely reflect our capabilities, it is a useful explanatory tool, and after reflecting on what *it* can do, I shall turn to show in what sense this idealization is relevant for explaining *our* use of the theory of statistical mechanics. Let us ask, then, how can Laplace's Demon predict whether S_0 will be in M_1 or M_2 at t_1 given that it started out in M_0 at t_0 and evolved subject to H.

The Demon, unlike us, knows the precise *actual micro*state of the universe at t_0, which is represented as a point in the $6N$-dimensional state space of the universe (three dimensions for the position and three for the velocity of each of the N particles in the universe). Knowing the Hamiltonian H (that is, the detailed equation of motion) of the universe, the Demon can also calculate its precise evolution and know its full microstate at each moment before or after t_0. *Ipso facto*, the Demon knows the microstate of S_0, which is a subsystem of the universe, and its evolution. The microstates of S_0 are represented as points in the $6K$-dimensional sub-space of the universe (three dimensions for the position and

three for the velocity of each of the K particles of S_0), and the possible evolutions of S_0 are projections of the $6N$-dimensional trajectory of the universe onto the $6K$-dimensional subspace of S_0. (I emphasize the distinction between the universe and S_0, and between the $6N$ space and the $6K$ subspace, since this distinction will be important for understanding a contemporary debate discussed later.)

Suppose now that we give Laplace's Demon the following task (see Figure 2). It should *pretend* (for a moment) that it does not know the full microstate of S_0 at t_0 and that all it knows about the initial microstate of S_0 at t_0 is that it has the aspect M_0 (so that in this respect, the Demon becomes a little bit like us, when we prepare a system with the property M_0). In terms of the $6K$-dimensional state space of S_0, all the Demon knows is that the actual microstate of S_0 at t_0 is somewhere in the region M_0. Based on this partial knowledge, it should predict the evolution of S_0.

Not knowing (by pretence) which microstate in the set M_0 is the actual one, the Demon carries out calculations for *all* the microstates in this set. It calculates, for each of these microstates, the trajectory segment that describes its evolution from time t_0 to time t_1. Call the resulting set of end points of all these trajectory segments $B(t_1)$; see Figure 2. Let us assume that some of the points of $B(t_1)$ are in M_1 and some are in M_2: In this case, the Demon will conclude that a system that starts out having aspect M_0 at t_0 will end up at t_1 with either aspect M_1 or aspect M_2.

That is good, but not enough: We would like to know what is the *transition probability* that the final microstate of S_0 (at t_1) will have the aspect M_1, given that it started out in M_0 at t_0 (and similarly for M_2). (In terms of the probability space in standard probability theory (Kolmogorov's, 1933), the microstates are the elements of the sample space, the sets Mi are the elements of the events space, and we are now looking for mechanical criteria for determining the probability function.) A reasonable answer would be this: Since we do not know which microstate in M_0 is actual, the transition probability is the fraction of the M_0 microstates that evolve to M_1 (and similarly to M_2). A reasonable way to determine this fraction is by looking at $B(t_1)$: The transition probability that the final microstate (at t_1) will have the aspect M_1 is given by the size of the overlap of $B(t_1)$ with region M_1, relative to the size of the entire region $B(t_1)$ (and similarly for M_2). But how should this relative size be determined? The difficulty is this: Since the state space is continuous, there are many measure functions that could be used to determine the size of its regions.[19] Which of these measures should the Demon use, and by what criterion? This question is under debate in contemporary literature. In this article, I shall not discuss this question and the various answers offered to it (these are addressed in Shenker, 2017) but only examine what *mechanics by itself* can provide towards answering it.

Figure 2 represents the case in which all that Laplace's Demon knows about the initial microstate of S_0 at t_0 is that it has the aspect M_0. But at this point, in order to choose a measure, we need to relax this constraint and allow the Demon to use some of its knowledge of the precise full microstate and trajectory of our actual universe and, *ipso facto*, of S_0. The idea is this: Use this Demonic *extended* knowledge in order to make physical sense of the probability statements made by creatures whose knowledge is *limited*. Here is how this can be done. Given the extended information, the Demon is asked to search in the entire actual universe—that is, along the *actual trajectory* of our entire universe—

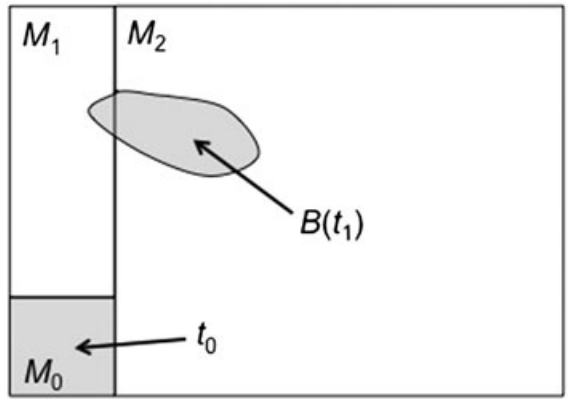

FIGURE 2 The emergence of probability in statistical mechanics

for systems that are similar to S_0 in all the respects that we deem relevant (call them S_1, S_2, etc.). (I should emphasize that the S_i systems are part of the *actual* universe, *not* in *counterfactual* universes. They are, of course, in different —and in that sense counterfactual— times and places.) For those S_i systems, the Demon is asked to do the following. Look along the (projected) trajectory of each of the S_i systems (in their respective $6K$-dimensional subspaces) for microstates that have the aspect M_0, follow the evolution of these microstates for a time interval equal to t_1–t_0, and notice the end points of those trajectory segments. In particular, we shall ask the Demon to notice how many of those end points have the aspect M_1 and how many M_2 (we know from the dynamics, illustrated in Figure 2, that these are the only options.); call these relative frequencies rf(M_1) and rf(M_2). (In this, the Demon becomes a bit more like us: We too can measure the relative frequencies in a number of systems similar to S_0 prepared with property M_0, although of course —unlike the Demon — not in all of them.) Finally, we shall ask the Demon to calculate which probability measures are such that, if imposed on the state space of system S_0, entail that the measure of the overlap of $B(t_1)$ with region M_1 is equal to rf(M_1), and the measure of overlap of $B(t_1)$ with region M_2 is equal to rf(M_2). In general, the result will not be unique, and in this case, the Demon will select one of these measures, either arbitrarily or according to some simplicity (or other) criterion. The resulting measure of overlap between $B(t_1)$ and each of the sets M_i will be the transition probability that we are after: It is the transition probability (or conditional probability) that the system will end up in either M_1 or M_2 at t_1 *given* that it started out in M_0 at t_0. (To be sure, the transition from relative frequencies to a probability measure involves an inductive leap, in the sense that the latter is applied also for microstates in the set M_0 that do not appear in our actual universe and were, therefore, not members of the set that the Demon had as input data. However, as far as this probability measure is used for predictions in the actual universe, it is not a generalization at all, but merely a convenient way of organizing the data, since the Demon measured *all* the M_0 cases that are on the trajectory of our actual universe. I address the counterfactual cases below.)

The *standard* approach to probability in statistical mechanics looks *prima facie* different, but in essence, it is embedded in the above scenario, and — importantly — by this embedding the above scenario explains the origin and status of probability in the standard presentations. I now present the standard view and describe its explanation along the above lines. *Suppose* that the chosen measure turns out to be the Lebesgue measure. (The Lebesgue measure is the natural generalization of the intuitive notion of size, and it is sometimes, in the context of the standard approaches, called the *standard measure*; e.g., in Albert, 2000). Since this measure is conserved under the dynamics (according to Liouville's theorem), so that the Lebesgue measure of $B(t_1)$ is the same as that of M_0, one may follow the trajectories backwards, as it were, from the $B(t_1)$ region to the initial set M_0, such that the region of overlap between $B(t_1)$ and M_1 is mapped to a subset of M_0 with the same Lebesgue measure, and similarly for M_2. In that case "the right probability distribution to use for making inferences about the past and the future is the one that's uniform, on the standard measure, over those regions of phase space which are compatible with whatever other information – either in the form of laws or in the form of contingent empirical facts – we happen to have [which in our case is that the initial microstate is in M_0]." (Albert, 2000, p. 96).

This statement is sometimes taken to be a *statistical postulate*, but as a postulate, this statement seems *mysterious* and its justification is unclear. In the above Demonic scenario, by contrast, since the Demon is a personification of the theory of mechanics, this statement is explained from mechanics by itself.

But what if, in the above Demonic scenario, the relative frequencies in our universe turn out to be such that the chosen measure is not the Lebesgue measure, but some other measure? In that case, the above story and the justification that it provides for the choice of probability measure in statistical mechanics remain intact: Carrying over the measure from $B(t_1)$ to M_0 would simply be more complicated. By contrast, if the Lebesgue measure is taken to be a "statistical postulate," then the usual reasoning for it would have to be abandoned. So in that sense, the above Demonic story is explanatory, whereas the "postulate" approach is not.

The conceivability that the chosen measure could turn out to be not the Lebesgue measure, but some other measure, emphasizes an important feature of our Demonic procedure: In that procedure, we have *described* the probabilities of magnitudes in our actual universe, not proven them from first mechanical principles. Attempts at such proofs

sometimes turn to arguing that the Lebesgue (or "standard") measure is "natural." Some of the arguments for the "naturalness" of the Lebesgue measure are based on its roles in theorems of mechanics (see, e.g., Werndl, 2013). The Lebesgue measure is conserved by the dynamics, according to Liouville's theorem, and is uniquely conserved if the dynamics is ergodic (in the sense of the Von Neumann–Birkhoff ergodicity theorem). However, these *diachronous* features of the Lebesgue measure do not entail that the *synchronous* relative size of the sets of initial conditions should be quantified by the same measure. Convenience and aesthetic considerations are not compelling.[20] (For more on whether the Lebesgue measure is "natural" in this context, see Hemmo & Shenker, 2015a, and Shenker, 2017)

The above derivation of transition probabilities is based on *idealizations* that call for justification: We need to show that the idealized Demonic story is relevant for explaining the success of statistical mechanical derivations using capabilities such as ours. One such idealization is the "demonic" calculation of the regions $B(t_1)$, M_1 and M_2, as well as the size of their overlaps according to the chosen measure. Needless to say, these calculations are "demonic" since there is no way we can carry them out, even approximately. However, first, this "demonic" story is an instrument to help us understand the meaning of probability statements in statistical mechanics, and in this sense, it is a useful and legitimate idealization. And second, there are practicable *shortcuts* that serve to replace the "demonic" calculations, which are based on reasonable and empirically verified assumptions (see, e.g., Hemmo & Shenker, 2012, Sections 6.5 and 6.6). One of these shortcuts is the hypothesis that the relations between the overlaps of $B(t_1)$ with region M_1 or region M_2 mirror the proportion between the Lebesgue measures of M_1 and M_2 themselves. Here, one would need the latter proportion, and the problem is that, again, the precise measures of the M_i regions are also unknown in practice. But there are reasonable approximations here too: for example, just assuming that the equilibrium set is overwhelmingly larger (by Lebesgue measure) than the rest suffices for many successful predictions. All these are *non-trivial* and certainly not a priori assumptions concerning the microstates of the system and its dynamics, but they are justified by being empirically successful in important cases. Importantly, realizing that these procedures *are shortcuts*, and justifying them *as such* (rather than taking them to be unexplained mysterious postulates), thus emphasizing that they are not a priori, helps us understand what is probability in statistical mechanics, how it is connected to observed relative frequencies, and how far we can build it from mechanics by itself.

Another idealization that goes into the above procedure for constructing the transition probabilities concerns calculating the relative frequencies with which all the similar S_i systems along the trajectory of our actual universe evolve into states with either property M_1 or property M_2 after a time interval of t_1–t_0. In practice we cannot, of course, carry out such a calculation, even approximately.

What can we do, with our limited capabilities, and how is it related to the above idealizations? The best we can do is start out by building many (I am not saying how many) systems that are identical to S_0 in all the parameters that we deem relevant and let them evolve during a time interval of t_1–t_0, all under the same conditions (that is, under the same Hamiltonian H, to a good approximation), making the very reasonable (but by no means a priori!) assumption that their initial microstates are different, and then measure whether each of them ends up at t_1 in M_1 or M_2. Call the relative frequencies of these end states rf(M_1) and rf(M_2).

Next, we imagine an abstract $6K$-dimensional subspace, which is not the space of any of our real systems S_i. Suppose we draw it as in Figure 2. Although—unlike Laplace's Demon—we cannot calculate the region $B(t_1)$, we are able to draw it *sketchily*, and only mark, on the partial overlap of $B(t_1)$ with M_1, the *observed* relative frequency of S_i systems that ended up at M_1 at t_1, and similarly for M_2. (To be sure, by doing so, we assume that all the systems S_i have the same number and type of degrees of freedom, which is never the case in the enormous and complex thermodynamic systems that we encounter in practice. This idealization, which can later be relaxed, is useful and provides explanatory insights.)

At this point, we make the following inductive leap, which is as acceptable as any generalization. On the basis of the little data that we have (a finite number of points in the $6K$-dimensional imaginary state space, based on measuring *some* of the M_0 systems in our world), we *conjecture* that the measured relative frequencies on other (yet unseen) M_0 systems in our world (other $6K$-dimensional subspaces of the actual universe) will be repeated in future measurements, and in this sense, we interpret these relative frequencies as probabilities (in a suitable sense of the term, which

I do not address here). To express this conjecture, we want to mimic what the Demon did in the above "Demonic" story, namely, we want to choose a *measure* over our imaginary *6K* state space so that the relative measures of the (sketchily represented) overlaps of $B(t_1)$ with M_1 and M_2 match these relative frequencies (and do so for the $B(t)$ of other moments as well, making sure that the same measure fits all of them). However, since (unlike the Demon) we do not have the details of the regions involved, we cannot carry out such a process. Instead, in practice, it is customary to *assume* that if we had those details, the result would be the Lebesgue measure. This choice is justified by successful predictions (not by any a priori considerations taking it to be "natural").

This is a reductive account of transition probabilities in statistical mechanics, that is, an explanation of how transition probabilities appear assuming *almost* only mechanics. The proviso "almost" refers to the fact that there is an extra element here, namely, extra measurements of relative frequencies that (presumably) did not go into formulating or supporting mechanics itself, and another extra element of generalizing these observations.

Remark: The above reasoning *cannot* be extended to explain why our universe is thermodynamic by taking it to be "typical" relative to counterfactual universes. Both the Demonic procedure and the "human" procedure described above are based on counting relative frequencies in our *actual* universe (by calculation in the Demonic case, by generalized observations in the human case). Since we cannot measure relative frequencies of the behavior of other universes (as we only have ontological access and *ipso facto* epistemological access to subsystems in our actual universe), this procedure is inapplicable for fixing a probability measure over the entire *6N*-dimensional space of the entire universe, and in particular over the initial conditions of the entire universe in its *6N*-dimensional space. We can perhaps imagine (for what it is worth) that the world could have started out in a different microstate than the one in which it actually started, but since we cannot make measurements on other worlds, the status of relative frequencies in this context is metaphysically precarious. (For further discussion of these issues see Hemmo & Shenker, 2015a, and Shenker, 2017.)

5 | CONCLUSION

In Section 1, I mentioned that as a point of *logic* time-asymmetric conclusions, such as the second law of thermodynamics, cannot be validly derived from time-symmetric assumptions, such as the laws of the fundamental theories of physics. The above description of what can be achieved with mechanics by itself, towards establishing the macroscopic and probabilistic terminology needed to explain the thermodynamic phenomena, inserted a temporal asymmetry by incorporating observed relative frequencies of transition *from* one state *to* another: in our example, from M_0 *to* either M_1 or M_2. This move is legitimate in the context of *describing* phenomena, not of *proving* them from first principles (a discussion of such insertions of temporal asymmetries is Price, 1996). If our aim is to *prove* from first principles *why* it is the case that we experience temporal asymmetry, on the basis of the temporally symmetric mechanical theory alone (without adding such state transition data), then we must (as a point of logic) add temporally asymmetric postulates to mechanics. For this *explanandum*, we need to add elements to the *explanans*. The postulates that are standardly added to mechanics in order to derive the thermodynamic phenomena concern measures, initial conditions, and Hamiltonians. The question is how to justify them. This question is addressed in Shenker (2017).

ENDNOTES

[1] See this characterization in, for example, Frigg (2008, p. 99) and Uffink (2007, p. 923). Naturally, once its tools are in place, the theory predicts other phenomena as well, which are justifiably seen as extensions of thermodynamics; see Wallace (2015) and Hemmo and Shenker (2012, 2015b).

[2] Howard (2010) takes this to be Einstein's most important contribution to the philosophy of science.

[3] A recent debate on the meaning of time symmetry in classical electrodynamics is described in Allori (2015) and references therein.

[4] These terms are not synonymous, see Hemmo and Shenker (2015a) and references therein.

[5] For the standard notions of "ideal gas" and "equilibrium," consult Frigg (2008). I touch upon the latter briefly below.

[6] See Wallace (2001, Section 1). See Ladyman & Ross, 2007 for a critical discussion of the use of outdated physics in philosophical discussions.

[7] For the quantum-mechanical counterpart of this idea, see Shenker (2018).

[8] Some have argued that applying statistical mechanics to systems with very few degrees of freedom is not legitimate, for example, Jauch and Baron (1972); see also Norton (2017). In this article, I clearly disagree with these arguments. See also Wallace (2015, p. 288).

[9] Constraints may decrease the number of degrees of freedom that are necessary for prediction.

[10] Their theories differ on the details of how the partial description yields predictions. For the differences between their approaches, see, for example, Callender (1999), Uffink (2007), and Frigg (2008).

[11] In this article, I am not committing myself to any particular meaning of "property"; see Orilia and Swoyer (2016).

[12] To make my point, I am ignoring the possibility of fluctuations. The Gibbsian conceptual framework is discussed below.

[13] In Hemmo and Shenker (2016), we use *macrovariable* to refer to any aspect of a microstate, and *macrostate* for those that appear in our experience.

[14] The argument for the H-theorem was found faulty; see Ehrenfest and Ehrenfest (1912), Uffink (2004), Brown et al., 2009.

[15] Such analogies became central in the approach of Gibbs, discussed below.

[16] In general, heat and temperature are different concepts, for example, one is intensive and the other is extensive. But they converge quantitatively in an ideal gas.

[17] For a more general discussion of the way in which different partial descriptions can yield different predictions, see Ben Menahem (2001).

[18] The non-triviality and the causal nature of such identity statements is even better seen by the history of their discovery; see Chang (2008).

[19] Measure being the mathematically precise generalization of the intuitive notion of "size."

[20] Another arguments for the "naturalness" of the Lebesgue measure in this context is based on the fact that it is used to quantify entropy (in both the Boltzmannian and the Gibbsian frameworks, albeit in different ways). However, the two notions are conceptually distinct, and conflating them may create the wrong impression that the probabilistic version of the law of approach to equilibrium is a tautology (see more on this point in Hemmo & Shenker, 2012, Section 7.3). Identifying probability and entropy also entails that the so-called past hypothesis postulates that the past state was improbable; see discussion in Callender (2004a, 2004b).

ORCID

Orly Shenker http://orcid.org/0000-0001-5716-6894

REFERENCES

Albert, D. (2000). *Time and chance*. Cambridge, MA: Harvard University Press.

Albert, D. (2014). *After physics*. Cambridge, MA: Harvard University Press.

Allori, V. (2015). Maxwell's paradox: The metaphysics of classical electrodynamics and its time-reversal invariance. *Analytica*, *1*, 1–19.

Ben Menahem, Y. (2001). Direction and description. *Studies in History and Philosophy of Modern Physics*, *32*, 621-635.

Brown, H., & Uffink, J. (2001). The origins of time-asymmetry in thermodynamics: The minus first law. *Studies in History and Philosophy of Modern Physics*, *32*(4), 525–538.

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

Callender, C. (1999). Reducing thermodynamics to statistical mechanics: The case of entropy. *Journal of Philosophy XCVI*, 348–373.

Callender, C. (2004a). Measures, explanation and the past: Should 'special' initial conditions be explained? *British Journal for the Philosophy of Science*, *55*, 195–217.

Callender, C. (2004b). Thermodynamic asymmetry in time. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2016 ed.). https://plato.stanford.edu/archives/win2016/entries/time-thermo/

Chang, H. (2008). *Inventing temperature: Measurement and scientific progress*. Oxford: Oxford University Press.

Ehrenfest, P., & Ehrenfest, T. (1912). *The conceptual foundations of the statistical approach in mechanics*. Leipzig: Name of publisher; reprinted 1990. New York: Dover.

Einstein, A. (1919). Time, space, and gravitation. Times London, 28 November 1919, 13–14. Reprinted as: What is the theory of relativity? In *Einstein: Ideas and opinions* (pp. 227–232). New York: Bonanza. 1954

Emch, G. (2007). Quantum statistical physics. In J. Butterfield & J. Earman (2006) (Eds.), *Philosophy of Physics* (pp. 1075–1182). Amsterdam: Elsevier.

Fermi, E. (1936). *Thermodynamics*. New York: Dover. 1956

Frigg, R. (2008). A field guide to recent work on the foundations of statistical mechanics. In D. Rickles (Ed.), *The Ashgate companion to contemporary philosophy of physics* (pp. 99–196). London: Ashgate.

Gibbs, J. W. (1902). *Elementary principles in statistical mechanics*. New Haven: Yale University Press.

Goldstein, S., & Lebowitz, J. (2004). On the (Boltzmann) entropy of nonequilibrium systems. *Physica D*, *193*, 53–66.

Gomori, M., Gyenis, B., & Hofer-Szabo, G. (2017). How macrostates come about. preprint available at http://philsci-archive.pitt.edu/id/eprint/12762

Hemmo, M., & Shenker, O. (2012). *The road to Maxwell's Demon*. Cambridge: Cambridge University Press.

Hemmo, M., & Shenker, O. (2015a). Probability and typicality in deterministic physics. *Erkenntnis*, *80*, 575–586.

Hemmo, M., & Shenker, O. (2015b). The emergence of macroscopic regularity. *Mind and Society*, *14*(2), 221–244.

Hemmo, M., & Shenker, O. (2016). Maxwell's Demon. In *Oxford handbooks online*Oxford University Press. view/10.1093/oxfordhb/9780199935314.001.0001/oxfordhb-9780199935314-e-63?rskey=plUl7T&result=1

Howard, D. A. (2010). Einstein's philosophy of science. *The Stanford encyclopedia of philosophy*. Summer 2010 Edition. Edward N. Zalta, ed. URL = http://plato.stanford.edu/archives/sum2010/entries/einstein-philscience/, Section 6.

Jauch, J. M., & Baron, J. G. (1972). Entropy, information and Szilard's paradox. *Helvetica Physica Acta*, *45*, 220–232.

Jaynes, E. (1965). Gibbs versus Boltzmann entropies. *American Journal of Physics*, *33*, 391–398.

Kolmogorov, A. N. (1933). *Foundations of the theory of probability*. English translation 1956. New York: Chelsea.

Kripke, S. (1980). *Naming and Necessity*. Harvard University Press.

Ladyman, J., & Ross, D. (2007). *Everything must go*. Oxford: Oxford University Press.

Landau, L. D., & Lifshitz, E. M. (1980). *Statistical Physics, Part 1, Course in Theoretical Physics, vol. 5* (3rd ed.) Trans: J. B. Sykes and M. J. Kearsley ed.). Oxford: Butterworth-Heinemann.

Lavis, D. (2005). Boltzmann and Gibbs: An attempted reconciliation. *Studies in History and Philosophy of Modern Physics*, *36*, 245–273.

Lavis, D. (2007). Boltzmann, Gibbs and the concept of equilibrium. *Philosophy of Science*, *75*, 682–696.

Lebowitz, J. (1993). *Boltzmann's entropy and time's arrow* (pp. 32–38). September: *Physics Today*.

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

Norton, J. (2017). The worst thought experiment. In M. T. Stuart, J. R. Brown, & Y. Fehige (Eds.), *The Routledge companion to thought experiments*. Oxford: Routledge.

Orilia, F., & Swoyer, C. (2016). Properties. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2016 ed.). https://plato.stanford.edu/archives/win2016/entries/properties/

Price, H. (1996). *Time's arrow and Archimedes' point: New directions for the physics of time*. New York: Oxford University Press.

Portides, D. (2017). Idealization and Abstraction in Scientific Modeling. Unpublished manuscript.

Ridderbos, K., & Redhead, M. (1998). The spin echo experiments and the second law of thermodynamics. *Foundations of Physics*, *28*(8), 1237–1270.

Shenker, O. (2014). Davidsonian Descriptions as a Principle Theory. *Iyyun: The Jerusalem Philosophical Quarterly*, *64*, 171–190. In Hebrew, translation to English available upon request.

Shenker, O. (2017). *Foundations of statistical mechanics: The auxiliary hypotheses*. Forthcoming: *Philosophy Compass*.

Shenker, O. (2018). Foundations of quantum statistical mechanics. In E. Knox, & A. Wilson (Eds.), *Companion to the philosophy of physics*. Oxford: Routledge. Forthcoming

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

Uffink, J. (2004). Boltzmann's work in statistical physics. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2008 ed.). http://plato.stanford.edu/archives/win2008/entries/statphys-Boltzmann/

Uffink, J. (2007). Compendium to the foundations of classical statistical physics. In J. Butterfield, & J. Earman (Eds.), *Handbook for the philosophy of physics, Part B* (pp. 923–1074). London: Elsevier.

Wallace, D. (2001). Implications of quantum theory in the foundations of statistical mechanics. Unpublished manuscript. Preprint available at http://philsci-archive.pitt.edu/410/

Wallace, D. (2015). The quantitative content of statistical mechanics. *Studies in History and Philosophy of Modern Physics*, *52*, 285–293.

Werndl, C. (2013). Justifying typicality measures of Boltzmannian statistical mechanics and dynamical systems. *Studies in History and Philosophy of Modern Physics*.

Werndl, C., & Frigg, R. (2015a). Rethinking Boltzmannian equilibrium. *Philosophy of Science*, *82*(5), 1224–1235.

Werndl, C., & Frigg, R. (2015b). Reconceptionalising equilibrium in Boltzmannian statistical mechanics. *Studies in History and Philosophy of Modern Physics*, *49*, 19–31.

Werndl, C., & Frigg, R. (2017). Mind the gap: Boltzmannian versus Gibbsian equilibrium. *Philosophy of Science*, *82*(5), 1224–1235.

Orly Shenker holds a PhD in philosophy of physics from the Hebrew University of Jerusalem, Israel, a BSc in physics from the same university, and an LLb from the Faculty of Law at Tel Aviv University, Israel. She is an associate professor at the Program for History and Philosophy of Science at the Hebrew University of Jerusalem, holds the Eleanor Roosevelt Chair in History and Philosophy of Science, and is the director of the Sidney M. Edelstein Centre for History and Philosophy of Science Technology and Medicine. She has co-authored (with Meir Hemmo) the book *The Road to Maxwell's Demon: Conceptual Foundations of Statistical Mechanics* (Cambridge University Press, 2012) and published papers on the foundations of classical and quantum statistical mechanics, the concept of probability, the foundations of quantum mechanics, and on physicalism in the special sciences and in the philosophy of mind.

How to cite this article: Shenker O. Foundation of statistical mechanics: Mechanics by itself. *Philosophy Compass*. 2017;e12465. https://doi.org/10.1111/phc3.12465