Two Approaches to Reduction: A Case Study from Statistical Mechanics

Abstract

I argue that there are two distinct approaches to understand reduction: the ontology-first approach and the theory-first approach. Further, I argue for the significance of this distinction by demonstrating that either one or the other approach has been taken as an implicit assumption in, and has in fact shaped, our understanding of what statistical mechanics is. More specifically, I argue that the Boltzmannian framework of statistical mechanics assumes and relies on the ontology-first approach, whereas the Gibbsian framework should assume the theory-first approach.

The relation between thermodynamics and statistical mechanics is one of the most paradigmatic instances of reduction. When one attempts to develop an account of reduction and needs an example to demonstrate how exactly that account works, the reduction of thermodynamics to statistical mechanics is the canonical case to which one appeals.[1] However, it is in fact questionable whether, and in what sense, thermodynamics can be reduced to statistical mechanics. Worse, it is not even clear what the correct theoretical framework or axiomatic foundation of statistical mechanics is: There are two competing frameworks for statistical mechanics on the table—the Boltzmannian and the Gibbsian frameworks,[2] and it is under contention which is correct.

Instead of assuming that we have a clear grasp of the reduction relation between thermodynamics and statistical mechanics and using that as a paradigmatic case to understand reduction, I propose to approach the problem from a different direction: I argue that there are two distinct approaches to understanding reduction—what I call *the ontology-first approach* and *the theory-first approach*. Furthermore, I argue that either one or the other approach has been taken as an implicit assumption in, and has in fact shaped, our understanding of what statistical mechanics is—in particular, whether its correct framework is Boltzmannian or Gibbsian.

To clarify, I don't intend to argue, in this paper, that either the ontology-first or the theory-first approach is the right approach to reduction. Rather, the point is to show why drawing a distinction between these two approaches is important and useful. How we understand reduction—either as ontology-first or theory-first—often is tacitly assumed and shapes our understanding of particular instances of reduction. (Interestingly, such assumptions can be so entrenched that some of those who incline toward either one of the approaches find it difficult to even understand the other

[1]See, for example, (Nagel 1961; Dizadji-Bahmani, Frigg, and Hartmann 2010; Schaffner 2012).

[2]This dichotomy can be found in, for example, (Callender 1999; Goldstein 2001; Frigg 2008; Wallace 2020). The more standard terminology is *the Boltzmannian approach* and *the Gibbsian approach* to statistical mechanics. I use "framework" instead to avoid confusion with the two approaches to reduction.

approach.) To demonstrate what role exactly the two approaches to reduction play, I turn to the reduction of thermodynamics to statistical mechanics as an example. In particular, given that a significant part (if not all) of statistical mechanics is to be a reductive underpinning of thermodynamics, these two approaches not only shape our understanding of the reduction relation in this instance, but also of statistical mechanics, the theory itself. I thus focus on statistical mechanics as an example.

In this paper, I first explicate the distinction between the ontology-first and the theory-first approaches to reduction. After setting up these two approaches in generality, I then move on to the example of statistical mechanics: In Section 2, I briefly introduce the essential elements of the Boltzmannian and the Gibbsian frameworks of statistical mechanics. In Section 3 and 4, I argue that the Boltzmannian framework, and especially the Boltzmannian criticisms of the Gibbsian framework, tacitly assume and rely on the ontology-first approach. In Section 5, I argue that the Gibbsian framework would be immune to these criticisms if it takes the theory-first approach as an assumption.

1 Ontology-first vs Theory-first Approach to Reduction

1.1 Introduction

Reduction is a relation. What are the relata of this relation? There is no univocal answer to this question.[3] Sometimes reduction is taken to be a relation between *objects* (that is, "real concrete things that exist here in our material world, things like quarks, or mice, or genes"[4]) at two dif-

[3]For a general review, see (van Gulick 2001, 3-4; van Riel and van Gulick 2019).

[4]To borrow from Cartwright (1983, 55).

ferent levels.[5] For example, a box of chlorine gas is composed of molecular chlorine; that is a reduction relation between the greenish yellow stuff in the box and chlorine molecules. Reduction of this kind is called *ontological reduction* or reduction as an ontological relation. (I use the term "*an* ontological relation of reduction" or "*a* reduction relation between objects" for a particular *instance* of ontological reduction.) Sometimes reduction is taken to be a relation between two scientific *theories*, for example, thermodynamics and statistical mechanics. Reduction of this kind is called *inter-theoretic reduction* or reduction as an inter-theoretic relation. (I use the term "*an* inter-theoretical relation of reduction" or "*a* reduction relation between theories" for a particular *instance* of inter-theoretic reduction.) Moreover, I intend to use the term 'ontological' in a broad way: to include reduction not just between objects, but also between their respective states, properties, quantities, etc. Crucially, though, ontological reductions relate properties that are not themselves especially theory-laden; that is, these properties can be understood independent of the relevant theories.[6]

How are these two kinds of reduction—ontological and inter-theoretic—related to one another? Is one more primary, on which the other depends? It may seem natural to think that inter-theoretic reduction is dependent on, and follows from, ontological reduction; that is, if objects at two levels bear a reduction relation (say, a composition relation between chlorine gas and molecular chlorine), then, as a consequence, the theories of the objects at each level should bear a reduction relation as well. But, does inter-theoretic reduction necessarily follow from ontological reduction? And what about the other way around? Despite the fact that both ontological and inter-theoretic reduction are commonly employed and discussed in various fields of philosophy and science, there has not been much explicit discussion of how these two kinds of reduction are, or should be, re-

---

[5]See, for example, (Smart 1959, 143; Ney 2013).

[6]The primary examples in this paper will be spatial locations of particles.

lated.[7]

This paper offers a starting point to think about the relation between ontological and inter-theoretic reduction. It identifies two competing ways to understand their relation: the ontology-first *approach* and the theory-first *approach* to reduction. These two approaches, in particular, are concerned with which kind of reduction is *prior to* the other.

1.2 An Account of Reduction: Ontology-first or Theory-first?

To clarify, neither the ontology-first approach nor the theory-first approach is meant to provide *an account of reduction*. An account of reduction concerns what reduction is. Usually, such an account forthrightly specifies what kind of reduction it is an account of (that is, whether its relata are objects or theories). It then identifies necessary and sufficient conditions that a successful reduction satisfies. Nagel's account (1961), which is probably the most well-known systematic account of reduction, takes the relata of reduction to be scientific theories. According to this account, one theory is reduced to another theory, if (roughly speaking) the former can be derived from the latter. Different accounts of reduction may identify different kinds of relata of reduction. Smart (1959, 143), to take another example, offers a tentative account that takes *entities* to be the relata of reduction.

In contrast, the two *approaches to reduction* concern the priority relation between ontological and inter-theoretic reduction, and can be conceived of as a way of classifying various accounts of reduction. By specifying what the relata of reduction are—whether they are objects or theories, a particular account of reduction takes reduction to be either *primarily* or *exclusively* an ontological relation (or an inter-theoretic relation). We thus can ask: For any given specific account of

[7] There are a few exceptions; see, for example, (van Riel 2014; van Riel and van Gulick 2019). McIntyre (2007) distinguishes the ontological and epistemological interpretations of reduction, but they are different from the ontology-first and the theory-first approaches.

reduction, does it follow the ontology-first approach or the theory-first approach?

To answer this question, we need to identify whether that account takes reduction to be *primarily* a relation between objects, or *primarily* a relation between theories. For example, Smart's account of reduction takes reduction to be primarily about entities, hence it is classified as following the ontology-first approach. Nagel's account, *prima facie*, may be seen as following the theory-first approach, since it takes reduction primarily to be a relation between theories.[8] More precisely, (i) if an account of reduction is committed to the idea that there is only one correct way to understand reduction, then we just need to identify whether it takes reduction to be *exclusively* an ontological relation or *exclusively* an inter-theoretic relation; (ii) if an account admits more than one correct way to understand reduction, then we need to identify whether that account takes ontological reduction to be prior (or primary) and inter-theoretic reduction to be derivative (or secondary), or the other way around.

What does it mean that ontological reduction is *prior to* inter-theoretic reduction (or the other way around)? Various senses of priority are adequate to flesh out the distinction between the two approaches to reduction. For instance, x is prior to y if y depends on, is derived from, is a consequence of, is justified by, or is explained by x. These different senses of priority are not mutually exclusive, but could be complementary. I will further exemplify the priority relation and

[8]Having said which, the status of bridge laws makes things more complicated. As noted by Sarkar (1992, 173) and van Riel and van Gulick (2019, 4.2), if bridge laws are conceived of as stating identities or relations between the extensions of terms in the reducing and the reduced theories, then "the characterization of the reductive link contains a metaphysical aspect," and "Nagel-reduction is a relation that holds not just between theories but also between their ontologies" (van Riel and van Gulick 2019, 2.2.3). Most importantly, "reduction on such a view incorporates essential reference to the theories' ontologies". That is to say, the ontological relations stated by the bridge laws are essential, and necessary, to reduction as an inter-theoretical relation. It thus seems to suggest that Nagel's account of inter-theoretic reduction requires ontological reduction after all. This, nonetheless, does not necessarily mean that Nagel's account follows the ontology-first approach, or the distinction between the ontology-first approach and the theory-first eventually collapses in this case. Instead, it can be seen as suggesting, for example, that: for Nagel's account to follow the theory-first approach (and thus to be compelling for those who believe that the theory-first approach is the right approach to reduction), bridge laws should not be understood as stating identities or relations between extensions.

the two approaches through specific examples.

An account of reduction follows the ontology-first approach if, for instance, it requires understanding ontological reduction to begin with in order to understand inter-theoretic reduction. Put in another way, what it is to be inter-theoretic reduction requires ontological reduction in the first place. For example, Oppenheim and Putnam's account of micro-reduction requires that the ontology of a branch of science "possess a decomposition into proper parts" of the ontology of another branch (1958, 6). Since they take this ontological reduction relation to be "the essential feature of a micro-reduction", their account follows the ontology-first approach. Moulines (1984, 55), to take another example, argues that a complete account of reduction between two theories requires ontological reduction between the respective domains. Moulines thus follows the ontology-first approach as well. In contrast, an account of reduction follows the theory-first approach, if ontological reduction is only a consequence of, and depends on, inter-theoretic reduction. New Wave Reduction is an account that most explicitly commits to the theory-first approach: It takes reduction to be primarily a relation between theories; more importantly, it is essential to this account that "the ontological consequences of a given reduction [that is, ontological reduction relations] *are secondary to and dependent upon* the nature of the theory reduction relation" (Bickle 1996, 65, 74).

Above, I briefly introduced the two approaches to reduction by demonstrating how an account of reduction can be classified as ontology-first or theory-first. Committing to an account of reduction is only one of the various ways for one to follow the ontology-first or the theory-first approach. I now explain what these two approaches are in more general terms.

1.3 *The Ontology-first* vs. *Theory-first Approach to Reduction: Further Explication*

The ontology-first approach takes it as given that there is a reduction relation between higher-level objects O_H and lower-level objects O_L, and *if there is a reduction relation between the theory of* O_H *and the theory of* O_L then this inter-theoretic reduction relation is a consequence of the ontological reduction relation. In short, the ontology-first approach takes ontological reduction to be *prior to* inter-theoretic reduction. This direction of priority is illustrated by Arrow (4) in Figure 1.

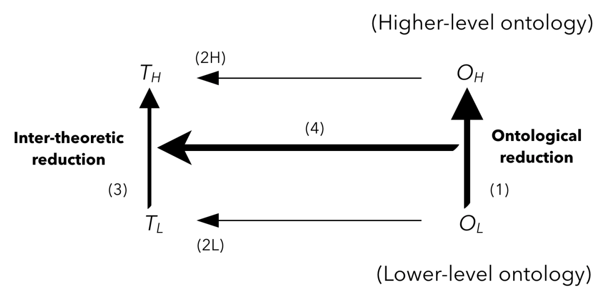

Figure 1: The ontology-first approach starts with the ontological reduction between O_H and O_L, illustrated by Arrow (1). Once O_H and O_L are specified, scientific theories are then meant to describe, explain, and make predictions about O_H and O_L—this direction from an ontology to its theory is illustrated by Arrow (2H) and Arrow (2L). Arrow (3) indicates that the theory of O_H reduces to the theory of O_L. Arrow (4) is the core of the ontology-first approach: it indicates that inter-theoretic reduction [Arrow (3)] follows from ontological reduction [Arrow (1)] as a consequence.

In contrast, the theory-first approach takes it as given that there is a reduction relation between two scientific theories T_H and T_L. Once T_H and T_L are each interpreted with an ontology, the approach states: *if there exists a reduction relation between the ontology of* T_H *and the ontology of* T_L, then this ontological reduction relation is a consequence of the inter-theoretic reduction relation. In short, the theory-first approach takes inter-theoretic reduction to be *prior to* ontological reduction.

This direction of priority is illustrated by Arrow (4) in Figure 2.

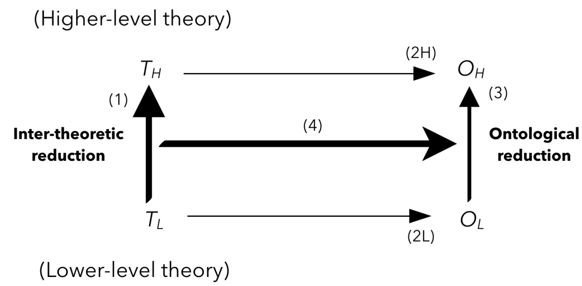

Figure 2: The theory-first approach starts with the inter-theoretic reduction between T_H and T_L, illustrated by Arrow (1). Given a theory, we can then interpret it with an ontology—this direction from a theory to its ontology is illustrated by Arrow (2H) and Arrow (2L). Arrow (3) indicates that the ontology of T_H reduces to the ontology of T_L. Arrow (4) is the core of the theory-first approach: it indicates that ontological reduction [Arrow (3)] follows from inter-theoretic reduction [Arrow (1)] as a consequence.

Stating these two approaches precisely requires specifying what ontological reduction and inter-theoretic reduction are, which requires specifying an account of reduction that takes objects as relata and another that takes theories as relata. However, neither the ontology-first approach nor the theory-first approach relies on any particular account of ontological reduction or inter-theoretic reduction. For our purposes, it suffices to get an intuitive idea of ontological reduction by thinking of, say, a mereological relation. An example of such a relation is the composition relation between chlorine gas and chlorine molecules (or, in mereological terms, chlorine molecules are parts of chlorine gas). Ontological reduction can also be understood in terms of supervenience, identity, realization, or elimination (van Gulick 2001, 4-9). For instance, the states of chlorine gas supervene on the states of chlorine molecules. The general idea of ontological reduction is that, as Schaffer (2008, 83) puts it, "[w]hat reduces is grounded in, based on, existent in virtue of, and nothing over and above, what it reduces to". He further uses a metaphor to illustrate this idea: "to create what reduces, God would only need to create what it reduces to" (Schaffer 2008, 83).

Crudely and tentatively, one can take inter-theoretic reduction to mean something like, one theory can be reduced to another if and only if the former can be fully explained by, or derived from, the latter. Consider, as a simplified example, reduction as *derivation*. In this case, what the theory-first approach takes as given are the two theories T_H and T_L, and, moreover, a derivation of T_H from T_L.

The main motivation behind the ontology-first approach is: Ontological reduction is about what the world is like, and what the world is like is independent of, and prior to, how we theorize about the world. Schaffer (2008, 83), for instance, expresses this idea: "Ontological reduction is independent of how we conceptualize entities, or theorize about them. Ontological reduction is a thesis about mind-and-theory-independent reality." Meanwhile, our scientific theories and any inter-theoretic relations, including inter-theoretic reduction, *depend on* what the world is like. This strongly suggests: inter-theoretic reduction depends on ontological reduction, not the other way around. Moreover, since ontological reduction means that O_H are "nothing over and above" O_L, the theories of O_H and O_L should also bear some kind of reduction relation—it would be deeply puzzling if they didn't. Fodor (1974, 97), for instance, has explicitly characterized this idea in the context of physicalism as "the assumption that the subject-matter of psychology is part of the subject-matter of physics is taken to imply that psychological theories must reduce to physical theories."

One (but not the only) way for ontological reduction to be prior to inter-theoretic reduction, or more generally how we theorize about objects, is if an ontology of a certain domain is taken to be prior to its theory. This way suggests two sufficient but not necessary conditions for an ontology-first approach:[9]

[9] I use these two conditions as a specific example to demonstrate what the ontology-first approach could be like, because they are important to our later discussion on how the ontology-first approach plays a role in Boltzmannian

First, scientific theories are primarily about objects. That is, given the objects from a certain domain, a scientific theory is meant to provide descriptions, predictions, and explanations of these objects. Hence, a theory, especially a physical theory, should forthrightly specify or postulate its ontology. Once what the ontology is has been made clear, only then does the theory say what the ontology does, how it behaves, that is, what its dynamics is.[10] A physical theory is thus necessarily attributed with an ontology. An uninterpreted mathematical formalism, even if it is successful at making novel predictions, does not count as a physical theory unless it is interpreted with an appropriate ontology.

Second, we can have some kind of grasp of what an ontology is like prior to any theory describing its behavior or dynamics. We may not know what exactly the ontology consists of, or what specific properties it possesses. Rather, what we can grasp are pre-theoretical or metaphysical constraints on what the ontology is like. That is to say, what the ontology of a theory is like is not only constrained by what is said by the theory, but also by pre-theoretical or metaphysical considerations. In particular, ontological reduction can be one of these considerations that act as a constraint on the ontology of a theory.

For example, Poidevin (2005) argues for the principle of recombination as a constraint on what chemical elements are physically possible, that is, on what the ontology of a chemical theory is. Elements in the periodic table (such as potassium [with atomic number 19] and calcium [20]), which form a discrete series, are *physically possible*. In contrast, anything with atomic number between 19 and 20 (say, 19.2 or 19.23), which forms a continuous series, is merely *logically possible*. According to the principle of recombination, the physical possibility of being a particular element is constituted by a recombination of actual instances of electron distributions (ibid., 129-130). Since there isn't any intermediate position between, say, having two electrons in one orbit around the

statistical mechanics.

[10]See Allori and Zanghi (2004, 1744), Maudlin (2010, 137; 2016, 318; 2019, 4), and Allori (2013, 63).

nucleus and having only one, anything with atomic number between 19 and 20 cannot be the result of a recombination of actual electron distributions and thus is ruled out as a physical possibility by the principle of recombination. This principle identifies a reduction relation between the higher-level objects, elements, and the lower-level objects, electrons, and it is this ontological reduction that determines what elements are physically possible and what are merely logically possible. Particularly, Poidevin (ibid., 131) emphasizes that the property of being a chemical element is theory-neutral. Hence, the reduction relation involved is indeed ontological rather than inter-theoretic.

Consider another example in which ontological reduction acts as a constraint on what the ontology of a theory could be and, consequently, on the theory itself. The primitive-ontology version of Bohmian mechanics has been defended by arguing that a fundamental physical theory (such as quantum mechanics) without a primitive ontology should be avoided. Primitive ontology was introduced as "the basic kinds of entities that are to be the building blocks of everything else" (Dürr, Goldstein, and Zanghì 1992, 850). Its role, as Allori (2015, 110) puts it, is to "ground a scheme of explanation" in which the behavior of the primitive ontology determines the properties of macroscopic physical objects. This suggests: introducing the primitive ontology secures a particular ontological reduction relation between the fundamental ontology and familiar higher-level objects (like tables, chairs, and measurement pointers). Given this relation between the fundamental ontology and familiar macroscopic objects and the latter being local and three-dimensional, the former needs to have these properties as well. It is in this way that ontological reduction imposes a constraint on what the ontology of the fundamental theory could be like; consequently, whatever the fundamental quantum theory turns out to be like, its ontology needs to contain the primitive ontology, or else it would not be the right theory.[11] [This argument would not work under the

[11] Or, at least, a fundamental theory with a primitive ontology should be preferred over those without one, because it is not clear or straightforward how the latter can give rise to familiar macroscopic objects.

The discussion in the literature has focused mostly on quantum mechanics (Dürr, Goldstein, and Zanghì 1992; Maudlin 2007, 2010), whereas the primitive-ontology approach proposed by Allori (2015) aims to be something more theory-first approach since the theory-first approach takes theories and their reduction relation to be primary but neither the theory that describes macroscopic objects like tables and pointers (namely, classical mechanics) nor its inter-theoretic reduction relation (with quantum mechanics) even comes up in this argument. This can be seen as an example of how the distinction between the ontology-first and the theory-first approach serves as an assumption in, and shapes, our understanding of a theory.][12]

In contrast to the ontology-first approach, the theory-first approach does not require each theory be attributed with an ontology. In other words, it is not necessary for a theory to forthrightly specify or postulate an appropriate ontology in order to be physical or carry any physical significance (instead of merely being a mathematical tool). How does a theory establish its status as a physical theory then? Via its usefulness or efficiency at describing patterns, making predictions, and providing explanations and practical applications. In physics, this is usually achieved by offering a new robust and autonomous dynamics (for example, Maxwell's equations offered such a dynamics for electromagnetic phenomena). Accordingly, a physical theory can be a mathematical formalism that is only partially interpreted, as long as it can be tested empirically, make novel predictions, and provide explanations.[13]

---

general.

[12]McCoy (2020b, 4, 10) also observes that the supporters of Bohmian mechanics and Boltzmannian statistical mechanics share the common ontological assumptions that these theories "are fundamentally about individual systems of microscopic entities" and "have a clear ontology of local beables"; moreover, they take this "ontological starting point as a point in their favor". McCoy himself argues against these ontological assumptions, and proposes his own interpretation of probability and statistical mechanics (2020a). His observation about Bohmian mechanics and Boltzmannian statistical mechanics intersects with my analysis in highlighting the ontological assumptions of these theories. McCoy, however, does not further explicate Bohmian-Boltzmannian justifications for such ontological assumptions. In my view, reduction is essential to these assumptions and their justifications. Moreover, McCoy's own view seems to me to assume the theory-first approach. Drawing the distinction between the ontology-first and the theory-first approach may help substantiate McCoy's position and observation.

[13]Rohrlich (1988, 303) sketches a view that takes the mathematical structure of the theory to be primary. For readers who are attracted to structural realism, one can think of the relation between the mathematical formalism (that is, a theory without an ontology) and the empirical world in terms of structural realism: the mathematical

Nevertheless, the fact that a theory is not necessarily attributed with an ontology does not imply that we cannot subsequently interpret the theory with an ontology. It's just that such an ontology does not play a primitive role in the theory. Anything that can be known about the ontology is given by the theory. Whether or not a theory is physical, or what its ontology is like, is not constrained by any metaphysical preconceptions about the ontology or the ontological reduction relation. Rather, the ontology can be seen only as secondary or derivative to its theory, especially to its dynamics. (This is easier to see with physical theories like quantum theories, less so with, say, biological theories.) A supplement on how we can attribute an ontology to a theory might be needed; for instance, something along the lines of functionalism or Dennett's (1991) pattern theory (I will say more about this in Section 5). The theory-first approach may demand a metaphysical picture that is radically different from what we are accustomed to: one no longer centered around objects with intrinsic properties moving in spacetime.

1.4 Clarificatory Remarks

The ontology-first and the theory-first approaches are not meant to exhaust the list of possible views on the relation between ontological and inter-theoretic reduction. They are better thought of as the two ends of a spectrum of views. One view on this spectrum, for example, is: ontological and inter-theoretic reduction are on a par, and if there is one of them, there is always the other (I will briefly mention some challenges to this view below). There might be another view which holds: there is no relation between ontological and inter-theoretic reduction. It's an odd possibility, and maybe an unlikely one. But it's not the purpose of this paper to defend or refute any of these particular views.

There are also variations within these two approaches to reduction. Recall that the ontology-

---

structure of the theory directly represents the world. Such a radical move, nevertheless, is not required by the theory-first approach.

first approach has a conditional, which leaves open the possibility that there isn't any inter-theoretic reduction following from ontological reduction. Thus, the view that ontological reduction is the only correct way to understand reduction or the only kind of reduction that holds in our world still counts as ontology-first. One possible reason why inter-theoretic reduction may not follow from ontological reduction: it is computationally intractable to derive the higher-level theory from the lower-level theory, even though the ontology of the former is reduced to the ontology of the latter.[14] Non-reductive physicalism is an example of an ontology-first view that is reductionist only about ontological reduction but not inter-theoretic reduction. Similarly, the view that there is only inter-theoretic reduction and no ontological reduction still counts as theory-first. For example, one may think that the ray theory of light is reduced to the wave theory, but there is no ontological reduction between the ray and the wave.

There are also various ways, as mentioned earlier, to understand the priority relation between ontological and inter-theoretic reduction. It might be helpful to think of the two approaches to reduction in terms of one more sense of priority: *explanatory priority*. In this case, the ontology-first approach states: ontological reduction *explains* inter-theoretic reduction; that is, the fact that O_H reduce to O_L (say, chlorine molecules are composed of chlorine atoms, etc.) explains the fact that the theory of O_H reduces to the theory of O_L (say, a chemical theory reduces to atomic physics). Similarly, the theory-first approach states: inter-theoretic reduction explains ontological reduction; that is, the fact that a theory T_H reduces to a theory T_L explains the fact that the ontology of T_H reduces to the ontology of T_L.

Nonetheless, neither approach suggests a chronological priority between ontological and inter-theoretic reduction (even if one understands the two approaches in terms of explanatory priority

---

[14] For more arguments on why inter-theoretic reduction might not follow from ontological reduction, see, for example, (Fodor 1974; List 2019).

and explanation as purely an epistemic notion[15]). Chronological priority would suggest (considering the theory-first approach as an example): we *first* find out, or construct, a derivation of one scientific theory from the other and thus discover a reduction relation between these two theories; *only then* would we know that there also exists a reduction relation between the ontologies of those two theories. The theory-first approach, on the contrary, does not require that we come to know the inter-theoretic relation of reduction first. In other words, this approach allows for the possibility that we may hypothesize an ontological relation of reduction, or even have a high credence in that hypothesis, before we know anything about the inter-theoretic relation of reduction. What the theory-first approach would say is that such a hypothesis is only justified in terms of the inter-theoretic relation of reduction, but not vice versa.

2 The Boltzmannian and the Gibbsian Frameworks of Statistical Mechanics

The Boltzmannian and the Gibbsian frameworks employ distinct concepts to describe the same physical systems, and there is no obvious way to translate the concepts of one to the concepts of the other (Frigg and Werndl 2019, 424). To present these two frameworks, let's consider again a box of chlorine gas. Suppose that the gas is initially confined in a corner of the box by a wall, and later spreads out to the whole box uniformly when the wall is removed. Suppose that the gas is composed of N chlorine molecules, a complete description of the microstate of the system at each time specifies the position q and momentum p of each molecule at that time.[16] The microstate thus can be represented by a point $(q_1, q_2, ..., q_N, p_1, p_2, ..., p_N)$ in the $6N$-dimensional phase space. This way of describing the system at the microscopic level is shared by the Boltzmannian and the Gibbsian frameworks. What they disagree on is how to describe the system *at the statistical-*

[15] See, for example, van Fraassen (1980) and Salmon (1984).

[16] For simplicity, we are assuming the system is classical and ignoring the internal degrees of freedom of the chlorine molecules.

mechanical level.

2.1 The Boltzmannian Framework

The Boltzmannian framework uses the concept of *macrostate* to describe the system at the statistical-mechanical level. A macrostate is characterized by macroscopic parameters, such as local pressure and local density of regions that are large enough to contain many molecules but small compared to the size of the box. It is related to the micro-description of the system, namely a microstate, as follows: The system in a particular macrostate could be in one of many different microstates, whereas the system in a particular microstate is in a unique macrostate. This is because what the macrostate of a system is is fully determined by the state of its microscopic constituents, but not vice versa. If we slightly change the location or velocity of just one particle in the system, the system would no longer be in the same microstate. But this change would not affect the local pressure or the local density of the system, or, more generally, its macrostate. Mathematically, the phase space can be partitioned into regions such that the microstates in each region correspond to the same macrostate of the system; a macrostate is thus identified with one of those regions. Regardless of whether it is the macrostate or the microstate that is under consideration, what is taken to be the object of study for the Boltzmannian framework is clearly, its advocates emphasize, *an individual system* (Frigg 2008; Goldstein 2019; Goldstein, Lebowitz, Tumulka, and Zanghi 2019).

Given the concept of *macrostate*, entropy and equilibrium are defined: The Boltzmann entropy of a system with macrostate M is

$$S_B \equiv k_B \ln \mu_M, \tag{1}$$

where k_B is the Boltzmann constant and μ_M is the phase-space volume of the macrostate M. For any given energy, there will be some macrostate which has the maximal Boltzmann entropy among all the macrostates with that energy. This state is designated as the equilibrium state in the Boltzmannian framework. As it turns out, the phase-space volume of the equilibrium state of a system at a given energy is overwhelmingly larger than any other macrostates with the same energy.[17] This feature of equilibrium is key to the Boltzmannian characterization of how systems approach equilibrium (such as how gas that is initially confined in a corner of a box will uniformly spread out to the whole box later) and their explanation of the *prima facie* inconsistency between the time-irreversibility of thermodynamics and the time-reversibility of its underlying micro-dynamics (i.e., classical mechanics).[18]

Many advocates of the Boltzmannian framework take the primary task of statistical mechanics to be to provide a microphysical description of and a justification of thermodynamics, and, in particular, to explain the time-irreversibility of thermodynamics.[19] Following some of Boltzmann's key insights, there has been a great deal of effort made to develop Boltzmannian statistical mechanics into a coherent and systematic framework,[20] and, relatively speaking, a consensus has been reached on what the Boltzmannian framework should be like.[21]

[17] For instance, for a system with $N \approx 10^{20}$, the ratio of the volume of an equilibrium macrostate to that of any non-equilibrium macrostate can be of order $10^{10^{20}}$ (Goldstein 2001, 43).

[18] See, for example, (Albert 2000; Goldstein 2001).

[19] See (Callender 1999; Wallace 2015).

[20] See, for example, (Callender 1999; Albert 2000; Goldstein 2001; Frigg 2008).

[21] Even though Boltzmann's own view on statistical mechanics went through various changes (Boltzmann 1872, 1895, 1896, 1897). Also see Uffink (2007). See Werndl and Frigg (2015) for an alternative Boltzmannian definition of equilibrium.

2.2 *The Criticized Gibbsian Framework*

Compared to the Boltzmannian framework, recent philosophy of physics has paid less attention to developing Gibbsian statistical mechanics into a systematic framework,[22] despite the fact that it is widely used among working physicists (Frigg and Werndl 2020) and is the standard tool in practical applications of statistical mechanics (Wallace 2020). Consequently, it is not clear what exactly the Gibbsian framework is (Frigg and Werndl 2020). For this reason, I first present the version of the Gibbsian framework that has been criticized by the Boltzmannians. Later in Section 4 and 5 I discuss possible conceptual modifications that can be made to the Gibbsian framework to respond to those criticisms.

In contrast to the object of study being an individual system in the Boltzmannian framework, the core object of study for the Gibbsian framework is commonly taken to be ensembles (Callender 1999; Frigg 2008; Goldstein 2019; Frigg and Werndl 2020) or probability distributions (Wallace 2020).[23] Gibbs's original characterization of an ensemble is as an infinite collection of systems of the same kind, which only differ in their configuration and velocities at a certain time.[24] Each system in this collection with its particular configuration and velocities is represented by a point in phase space. The state of the ensemble at time t is represented by a probability density function $\rho(q, p; t)$ over the phase space. The time evolution of ρ is given by Liouville's equation:

$$\frac{\partial \rho}{\partial t} = -\{\rho, H\}, \tag{2}$$

where H is the Hamiltonian and $\{\ ,\ \}$ is the Poisson bracket.

[22] For such attempts, see (Malament and Zabell 1980; Sklar 1993; Wallace 2020).

[23] Talk of *ensemble* is prevalent in physics. However, it is ambiguous whether physicists take ensembles to be literally the object of study, or just a heuristic to talk about probability distributions. More on this in Section 4.

[24] For illustrative purposes, we are again working with the example of a box of chlorine gas.

In the Gibbsian framework, the entropy of a system is first defined as

$$S_G(\rho) \equiv -k_B \int_\Gamma \rho \ln(\rho) d\Gamma, \tag{3}$$

where Γ is the phase space and $d\Gamma$ is the standard Lebesgue measure. This is the so-called Gibbs fine-grained entropy. It is invariant over time, as a consequence of Liouville's equation. Since thermodynamic entropy increases when the system evolves from a non-equilibrium state towards an equilibrium state, the Gibbs fine-grained entropy is inadequate to be the microphysical counterpart of thermodynamic entropy. A notion of entropy that is not invariant in time, the Gibbs coarse-grained entropy, is thus introduced.

Abstractly, coarse-graining is a procedure of averaging over details of the system that are irrelevant to the description of the system at a higher level. We can represent such a procedure by a projection operator J, which is a map on the space of probability distributions such that $J^2 = J$ (i.e., the result of coarse-graining twice is the same as coarse-graining once). J acts on the original probability density ρ, yielding the coarse-grained density:[25]

$$\bar{\rho} = J\rho. \tag{4}$$

One particularly important way, at least conceptually, to think of coarse-graining is as partitioning the phase space into small cells. We then define a coarse-grained density $\bar{\rho}$ such that it is uniform over each of the cells and assigns the same probability to a cell as the original probability density ρ. $\bar{\rho}$ is coarse-grained in the sense that the details of ρ within each cell are disregarded.

In the Gibbsian framework, the microphysical counterpart of thermodynamic entropy is the Gibbs coarse-grained entropy $\bar{S}_G$. It has the same form as Eq.3, but substitutes ρ with the coarse-

[25]This method is usually referred as *the method of projections*. For more details, see (Zwanzig 1960, 1966; Wallace 2015, 2020; Robertson 2018).

grained density $\bar{\rho}$:

$$\bar{S}_G(\rho) \equiv S_G(\bar{\rho}) = -k_B \int_\Gamma \bar{\rho} \ln(\bar{\rho}) d\Gamma. \tag{5}$$

Accordingly, equilibrium is defined as a state for which $\bar{\rho}$ is invariant in time. The Gibbsian framework characterizes how systems approach equilibrium in terms of the increase of the Gibbs coarse-grained entropy. To describe and make quantitive predictions about thermodynamic systems at equilibrium, the Gibbsian framework associates each macroscopic parameter with a phase function $f : \Gamma \to \mathbb{R}$. The phase average $\langle f \rangle$ of f:

$$\langle f \rangle = \int_\Gamma f(q,p)\rho(q,p;t)d\Gamma \tag{6}$$

gives the values of these macroscopic parameters.[26] Precisely because the macroscopic parameters are insensitive to coarse-graining, we in fact attain the same value for $\langle f \rangle$ whether we use the original ρ or the coarse-grained $\bar{\rho}$.

To summarize, the Boltzmannian and the Gibbsian frameworks offer different descriptions of the same physical system at the statistical-mechanical level; in particular, they differ in whether such descriptions should involve probability. For the Gibbsian framework, probability (or ensemble) is indispensable to represent the system and to define key notions like entropy and equilibrium. For the Boltzmannian framework, it's not.[27]

---

[26]This is the standard way to calculate equilibrium thermodynamic values used by working physicists. In Boltzmannian framework, equilibrium thermodynamic values are just macroscopic values which specify macrostates (Wallace 2020).

[27]To clarify, this issue—whether we can use probability to represent systems and to define entropy and equilibrium—is different from a more general issue on whether statistical mechanics should use probability at all. On the latter, some defenders of the Boltzmannian framework argue that probability should be replaced by typicality (Goldstein 2012), whereas some others do not think the use of probability can be completely eliminated (for instance, some think probability is needed to characterize the initial condition of the system [Albert 2000]). There are also efforts made to accommodate probability in the Boltzmannian framework (see, for example, Frigg and Hoefer (2015)).

3 The Ontology-first Approach to Reduction and the Boltzmannian Framework

In this section, I argue that the Boltzmannian framework of statistical mechanics assumes and relies on the ontology-first approach. To clarify, I do not mean to argue that the advocates of the Boltzmannian framework just happen to assume the ontology-first approach. Nor do I mean to argue for the historical claim that Boltzmann or his followers had the ontology-first approach as an assumption in mind while developing the theory. I do not intend to argue for a logical claim either: It is not the case that the Boltzmannian framework is entailed by the ontology-first approach (plus some other assumptions). What I mean is something conceptual: in order to make sense of the Boltzmannian framework, we need to assume the ontology-first approach.

To show this, it needs to be shown how the framework directly appeals to ontological reduction, more specifically, an ontological relation of reduction that holds between thermodynamic and statistical-mechanical systems. If it were the case that the Boltzmannian framework instead assumed the theory-first approach, then the ontological relation of reduction would be secondary or derivative and thus would not appear directly in the framework.

Recall how the key concept in the Boltzmannian framework, *macrostate*, is related to *microstate*: a microstate corresponds to a unique macrostate, while a macrostate is compatible with many different microstates. How is this relation justified? The obvious justification appeals to ontological reduction. It is because of the ontological relation of reduction (say, the mereological relation between chlorine gas and chlorine molecules) that a microstate of the molecules and the corresponding macrostate of the gas are just two descriptions of the same system and that these two descriptions are related in this particular way. If an ontological relation of reduction were not assumed, the fact that there is a relation between macrostate and microstate would not be natural and obvious, and we would request some other justification as to why macrostate and microstate are related in this particular way. But no such request has been made.

Instead, supporters of the Boltzmannian framework are explicit that an ontological relation of reduction is taken to be an assumption in their discussions of inter-theoretic reduction between thermodynamics and statistical mechanics. For example, Callender (1999, 366) claims:

> We know that . . . the actual gas has a microstate X. We also know that X, whatever it is, gives rise to the macrostate M we see before us. These are merely the assumptions we make when we say thermodynamics is in some sense reducible to mechanics. They are completely uncontroversial. Surely, the gas has a microstate, and surely whatever microstate it occupies corresponds to the macrostate we see.[28]

Moreover, Callender distinguishes ontological reduction from inter-theoretic reduction. He thinks only the latter poses a real problem for reducing thermodynamics to statistical mechanics, whereas it is an uncontroversial assumption that thermodynamic systems are "ontologically reduced" to mechanical systems (ibid., 351):

> Thermodynamic systems—like chairs, tables, and similar systems picked out by our common object language—are nothing more than complicated arrangements of physical properties. Very few would disagree with this. Thermodynamics does not threaten physicalism. In this weak sense, thermodynamics is already 'ontologically reduced' to mechanics.

Frigg (2008, 104), to take another example, points out that reduction between a macrostate and a microstate is taken to be an assumption in the Boltzmannian framework, and he characterizes this ontological relation of reduction in terms of supervenience:

---

[28]Although the context of this claim is to discuss a problem for the Gibbsian framework, it does not make a difference for our purposes. Because Callender takes this claim to hold both in the Boltzmannian and the Gibbsian frameworks.

> It is one of the basic posits of the Boltzmann approach that a system's macro-state supervenes on its fine-grained micro-state, meaning that a change in the macro-state must be accompanied by a change in the fine-grained micro-state.

One concern may be raised against this argument. Even if the Boltzmannian framework assumes ontological reduction, it does not mean that the framework assumes the ontology-first approach. That the theory-first approach takes inter-theoretic reduction to be prior to ontological reduction does not mean the approach is incompatible with there being an ontological relation of reduction. It may well be the case that (a) the Boltzmannian framework assumes the theory-first approach and an ontological relation of reduction at the same time, but that the ontological reduction is only secondary to or derivative from an inter-theoretic relation of reduction. Or (b) the Boltzmannian framework assumes both ontological and inter-theoretic reduction and takes them to be on a par (that is, neither is prior to the other).

Both (a) and (b) are possible, but not plausible. If (a) were true, the role of ontological reduction in the Boltzmannian framework could thus be fulfilled by some kind of inter-theoretic reduction. That is to say, it should at least be possible for the framework to be presented or reformulated in a way that does not directly appeal to ontological reduction. However, this is not how the Boltzmannian framework is presented, and it is not straightforward how this can be done.[29] If the theory-first approach were assumed, a more straightforward justification of the relation between macrostate and microstate would appeal to, say, how the *dynamics* at the macro-level is related to the *dynamics* at the micro-level. However, it is unclear, or at least not obvious, how the relation between macrostate and microstate can be explained in terms of the dynamics of the system in

[29] I do not mean that the Boltzmannian framework necessarily assumes the ontology-first approach, or that there is some intrinsic feature of the Boltzmannian framework that makes it impossible for it to be adapted to the theory-first approach. It may well be possible that the framework can be reformulated in a way that does not assume the ontology-first approach. I mean to argue, rather, that the Boltzmannian framework, as presented the way it is, assumes the ontology-first approach.

the Boltzmannian framework. If anything, it goes the other way around: The Boltmannian derivations or justifications of the second law or the dynamical equations of thermodynamics assume the relation between macrostate and microstate. That is to say, inter-theoretic reduction in the Boltzmannian framework is not primitive but something derived. Hence it cannot be on a par with ontological reduction, given that the latter is assumed by the framework.

4 The Ontology-first Approach to Reduction and the Boltzmannian Criticisms of the Gibbsian Framework

The role of the ontology-first approach is even more explicit in Boltzmannian criticisms of the Gibbsian framework.

4.1 The Problems of Ensemble and Probability

First of all, the Boltzmannians often criticize the Gibbsian framework for taking "ensembles of infinitely many systems" as its core object of study. In particular, they criticize it for using ensembles to represent *actual individual* systems (Callender 1999; Albert 2000; Goldstein, Lebowitz, Tumulka, and Zanghi 2019). In fact, the Boltzmannians describe the Gibbsian framework as "ensemblist", in contrast to their own framework being "individualist" (Goldstein 2019). The criticism goes as follows. Statistical mechanics should be about *actual individual* physical systems. An ensemble, which is a collection of infinitely many systems, is neither actual nor individual. More specifically, equilibrium and entropy are supposed to be properties of an individual system. But if they are defined in terms of probability distributions over ensembles, then an individual system can no longer be said to be in equilibrium or have certain entropy. Moreover, we cannot infer the behavior of an individual system from the behavior of an ensemble (Frigg 2008, 174). Therefore, actual individual systems cannot be represented by ensembles.

The Gibbsians have an immediate response to this criticism: An ensemble is only a fictitious set of all possible microstates of the system. It is introduced merely for convenience or as a heuristic. In fact, the Gibbsian framework can be presented without mentioning 'ensemble' at all: Statistical-mechanical systems are directly represented by probability distributions over phase space.[30]

But this is criticized by the Boltzmannians as well. They argue that *actual*, *individual*, statistical-mechanical systems cannot be represented by probability distributions. Here's Callender (2001, 544), for example:

> The problem is not the use of ensembles . . . The problem is instead thinking that one is *explaining* the thermal behaviour of *individual real systems* by appealing to the monotonic feature of some function, be it [of] ensembles or not, that is not a function of the dynamical variables of real individual systems.
>
> It is impossible to calculate the intellectual cost this mistake has had on the foundations of statistical mechanics. (Emphasis Callender's.)

For Callender, any function that is not "a function of the dynamical variables of real individual systems" is inadequate to be a part of the explanation for the thermal behaviors of actual individual systems, and probability is one such function.

Why can't a probability distribution represent "individual real systems"? A probability distribution describes how likely it is for a system to be at one of the many possible microstates. But, at any given time, there is only one definite microstate at which the actual system can be. Goldstein (2019, 443) thus asks:

> What, after all, does *the* probability distribution μ_t of our system at a given time

[30] For a presentation of the Gibbsian framework without ensembles, see Wallace (2020).

> refer to? What in fact is its actual probability distribution? I'm aware of no plausible answer to this question.

By pointing out that a non-trivial probability distribution over many possible microstates does not refer to anything actual, Goldstein is effectively arguing that it is problematic to use probability to represent an actual system.

(One may argue that an actual individual system can be represented by probability, if probability is interpreted as subjective in the sense that it measures how much we know about the system. In that case, Gibbs entropy, which is defined in terms of probability, would be subjective as well. However, the reason why thermodynamic entropy of an isolated system does not decrease cannot be subjective, since that fact holds regardless of how much we know about the system. Thus, Gibbs entropy as a subjective notion is not adequate to capture this objective fact about thermodynamic entropy.[31] Accordingly, interpreting probability as subjective is not a viable solution to the problem of whether probability can represent actual individual systems.)

The key to the Boltzmannian criticisms of the Gibbsian use of ensemble and probability lies in their claim that statistical mechanics should be about *actual individual systems*. If a framework of statistical mechanics is not about individual systems for whatever reason, it is plainly a drawback of that framework. Here's Callender (1999, 357) again:

> Thermodynamics states that once an isolated system achieves equilibrium, it stays in equilibrium forever . . . Boltzmannian SM...abandons the idea that equilibrium is stationary in time. The Boltzmann approach balances this affront to thermodynamics by retaining the idea that equilibrium and entropy are properties of *individual systems*. The Gibbs approach *pays for* its strict agreement with the thermodynamic laws

[31] For more details, see (Albert 2000; Goldstein, Lebowitz, Tumulka, and Zanghi 2019).

> by relinquishing the idea that entropy and equilibrium are properties of individual systems. (Emphasis mine)

For Albert (2000, 70), it is just "sheer madness" that entropy, equilibrium, and the laws of thermodynamics are associated not with individual systems, but with ensemble or probability.

Why is statistical mechanics supposed to be about actual individual systems? We can answer this if we take scientific theories to be primarily about objects (see Section 1.3). Then it is reasonable to think that appealing to objects and their actual states is the only admissible way to characterize a given physical system, and appealing to "some abstract entity", such as probability (Maudlin 1995, 147), is not. Accordingly, the core object of study of a physical theory cannot be a fictitious collection of many possible states. Since taking theories to be primarily about objects is a sufficient condition for the ontology-first approach (see Section 1.3), this approach needs to be assumed as a consequence. In sum, to justify their claim that statistical mechanics is supposed to be about actual individual systems in this way, the Boltzmannians assume the ontology-first approach.

Maudlin (1995, 147) gives a slightly different explanation, which arguably appeals to ontological reduction:

> Since phenomenological thermodynamics originally was about such individual boxes [of gas before me], about their pressures and volumes and temperatures, 'saving' it by making it be about probability distributions over ensembles seems a Pyrrhic victory.

That is to say: since thermodynamics is about individual systems, statistical mechanics is supposed to be about individual systems as well; making statistical mechanics be about probability distributions, even though it has the benefit of preserving thermodynamics, has the cost of making statistical mechanics no longer be about individual systems; this cost is so devastating that it is tantamount to defeat. But why is *statistical mechanics* supposed to be about individual systems?

Simply because thermodynamics is about individual systems? If the ontology-first approach is assumed, then ontological reduction acts as a constraint on what the core object of study of a theory could be (see Section 1.3). In this case, given that there is an ontological relation of reduction between thermodynamic and statistical-mechanical systems and that thermodynamic systems are individual systems, statistical-mechanical systems should be individual systems as well.

4.2 The Problem of Coarse-graining

Moreover, the Gibbsian framework is criticized by the Boltzmannians for its use of coarse-graining without proper justification. For instance, Callender (1999, 360) argues that the sole purpose of coarse-graining is to get a notion of Gibbs entropy that monotonically increases in time (i.e., the Gibbs coarse-grained entropy; see Eq. 5). That is to say, the coarse-graining projection operator J is chosen opportunistically, and thus the Gibbs coarse-grained entropy is introduced without any justification apart from it matching the increase of thermodynamic entropy in time.

The Boltzmannian framework, however, also employs coarse-graining without providing a justification. Although the Boltzmannians do not always use the word "coarse-graining", the idea of partitioning the phase space into finite regions is explicit in the framework. For example, the definition of Boltzmann entropy appeals to coarse-graining:

> [W]e define the Boltzmann entropy S_B for the actual microstate of an individual system. Consider some microstate X. X corresponds to a macrostate $M(X)$, which, in turn, is compatible with many different microstates. We wish to determine the relative volume in [the phase space] corresponding to all the microstates giving rise to M. To accomplish this, we must partition [the phase space] into compartments such that all of the microstates X in a compartment are macroscopically indistinguishable. (Callender 1999, 355)

Hence, the Boltzmannians apply a double standard in criticizing the Gibbsian use of coarse-graining.[32] What justifies this double standard? The most plausible answer I can think of is that the Boltzmannian framework tacitly assumes the ontology-first approach. This assumption licenses it to take ontological reduction as given, which justifies its choice of coarse-graining, more specifically, its choice of partitioning the phase space into "macroscopically small but microscopically large cells" (Goldstein 2001, 42). This particular choice of the size of the cells can be justified, as Frigg (2008, 135) points out, if there exists an objective separation of the relevant macroscopic and microscopic scales. A reduction relation between objects at a microscopic and a macroscopic scale provides a natural and objective micro-macro separation. It's simply "carving nature at its joints" that there are such objects at such and such scales. The two scales involved in coarse-graining are thus not chosen arbitrarily, but are picked out because they are ontologically significant (that is, they are associated with the relevant ontologies in the relevant reduction relation). Put in another way, the choice of coarse-graining is justified since it gives rise to the right higher-level objects, which is marked off by what is macroscopically indistinguishable and what can be meaningfully measured. This seems to be the kind of justification that Albert (2000, 42) alludes to when he says "Everyday macroscopic human language . . . carves the phase space of the universe into chunks", right before he introduces coarse-graining. On the other hand, without the assumption of an ontological reduction relation, it would be puzzling why being macroscopically indistinguishable even matters to any individual microstate.

The Gibbsian framework, however, cannot appeal to the same kind of justification. In this framework systems are represented by probability distributions. It is not clear in what sense probability distributions are ontological, and stand in (or can stand in) an ontological reduction relation with individual microstates. Indeed, this is exactly what the Gibbsian framework is criticized for (under the assumption of the ontology-first approach, as discussed in Section 4.1). Hence,

[32]One exception is Frigg (2008, 134-135).

the Boltzmannians would argue, there is no ontological reduction for the Gibbsian framework to appeal to to justify their choice of coarse-graining.

5 The Theory-first Approach to Reduction and the Gibbsian Framework

It is subtler how the Gibbsian framework is related to the theory-first approach. Unlike the case of consciousness or biological phenomena, the ontological reduction relation between thermodynamic and statistical-mechanical systems is relatively clear and uncontroversial. (The Boltzmannian vs the Gibbsian debate is not a debate about whether, say, a box of gas is composed of molecules.) The ontology-first approach thus appears to be the prevailing assumption in discussions of statistical mechanics, even among those that are more on the side of the Gibbsian. For instance, Malament and Zabell (1980, 341) claim:

> Every one of these [thermodynamic parameters], presumably, is uniquely determined by the exact microstate of the gas. That is our fundamental reductionist assumption.

This quote suggests that the ontological reduction relation between thermodynamic and statistical-mechanical systems has been assumed even in a Gibbsian discussion (even though it is unclear what exact role this assumption plays in their argument). Thus, I do not intend to argue that the Gibbsian framework assumes and relies on the theory-first approach. Rather, I argue that the Gibbsian framework would be immune to certain criticisms if it assumes the theory-first approach. In fact, the Gibbsian framework is vulnerable to those criticisms discussed earlier exactly because the criticisms are taken from the point of view of the ontology-first approach. Hence, for the Gibbsian framework to be a coherent and valid foundation for statistical mechanics, it should assume the theory-first approach.

The main reason for the Gibbsian framework to adopt the theory-first approach is: this approach permits the use of probability to represent statistical-mechanical systems. As mentioned in Section 2.2, Gibbsian statistical mechanics is the standard tool used among working physicists. It is more efficient than, say, classical mechanics at describing patterns, making predictions, and providing explanations in certain domains. If the theory-first approach is assumed, then the efficiency and usefulness of the Gibbsian framework warrant its status as a viable and physically significant theory. Accordingly, what the theory's core object of study is, or how a system can be represented, is not constrained by any pre-theoretical or metaphysical considerations (see Section 1.3). Consequently the Gibbsian framework cannot be ruled out as a viable physical theory *just because* it does not describe the appropriate ontology. Put in another way: without the ontology-first approach, it is unclear why systems cannot be represented by probability distributions (as discussed in Section 4.1).

To clarify, this argument does not in any way suggest that the Gibbsian framework conflicts with the presence of an ontological relation of reduction, such as the composition relation between a box of gas and its molecules. Recall: according to the theory-first approach, ontological reduction (if there is any) follows from inter-theoretic reduction. Hence this approach is not incompatible with there being an ontological reduction relation; it's just that such ontological reduction should be conceived of as secondary to inter-theoretic reduction. This can be illustrated if we consider a particular way in which probability can represent statistical-mechanical systems. If we adopt something along the lines of Dennett's pattern theory (1991), the representational role of probability can be justified in terms of its efficiency in describing the dynamical patterns picked out by Gibbsian statistical mechanics. What Dennett's pattern theory contributes is to explain and justify how a new higher-level ontology can emerge in terms of some real patterns,[33] even if such an ontology is not reduced to a lower-level ontology in a way that we are familiar with (such as in terms of composition or identity).

[33] Roughly speaking, there is a real pattern if and only if there is some more efficient way to describe a certain phenomenon than specifying every single detail about it.

This way for probability, or whatever it represents, to emerge as a higher-level ontology is unsuited for the ontology-first approach. Because it is essential to this approach that any ontological reduction (including the one probability stands in) is taken to be primary. If there is an ontological reduction relation between probability and lower-level objects like molecules, it would not be a standard one (like composition), especially since probability distributions are not located in ordinary physical space. Such a new ontological reduction needs to be either postulated as primitive or justified. It's hard to see how the former can be motivated. The latter would require the theory of statistical mechanics (in particular, its dynamics involving probability) and the underlying mechanics. That is to say, the ontological reduction relation in which probability stands (if there is any) is dependent on and secondary to the inter-theoretic reduction relation.

The second reason for the Gibbsian framework to adopt the theory-first approach is that it provides a way for the framework to justify its choice of coarse-graining. Again it is according to the theory-first approach that Gibbsian statistical mechanics can establish its status as a physical theory via its usefulness and efficiency, and a typical way for a physical theory to have these virtues is to identify a new robust and autonomous dynamics (see Section 1.3). In our case, an appropriately chosen coarse-graining procedure can give rise to such a new robust and autonomous dynamics for the relevant degrees of freedom (Wallace 2015, 2020; Robertson 2018). The coarse-graining projection operator J decomposes the original probability ρ into a relevant part $\bar{\rho} = J\rho$ and an irrelevant part $\bar{\rho}_{irr} = (1 - J)\rho$. What is special about this decomposition is that there turns out to be autonomous dynamical equations for $\bar{\rho}$ (namely, the coarse-grained probability). In a sense, J throws away the part of the original probability distribution that is irrelevant to the new dynamics. The existence of such dynamics thus justifies the particular choice of coarse-graining. Accordingly, contrary to what the Boltzmannian critics think, the coarse-graining is not chosen in the Gibbsian framework *just* to match the increase of thermodynamic entropy.

Since the dynamics of $\bar{\rho}$ is obtained from the corresponding theory (i.e., Gibbsian statistical mechanics), this justification of coarse-graining appeals to inter-theoretic reduction, instead of ontological reduction. Recall the earlier discussion on probability: if there is any ontological reduction in which $\bar{\rho}$ stands in, it would be dependent on and secondary to the inter-theoretic reduction relation. This is why this justification of coarse-graining would not work under the ontology-first approach. Worse, the ontology-first approach would question if the dynamics of $\bar{\rho}$ is even physical, since $\bar{\rho}$ does not move in ordinary physical space.[34] In contrast, the theory-first approach allows for the emergence of a new dynamics (that is, the time evolution of $\bar{\rho}$, a real pattern) and the introduction of a new higher-level ontology picked up by that dynamics.

Lastly, adopting the theory-first approach permits a broader understanding of statistical mechanics than is conceived of by the ontology-first approach, and this broader understanding is more congenial to the Gibbsian framework than to the Boltzmannian framework. As mentioned in Section 2.1, the primary task of the Boltzmannian framework is taken to be to provide a microphysical description and a justification of thermodynamics. In contrast, the scope of the Gibbsian framework goes beyond that (Wallace 2015): it contains a collection of techniques that are used to model all kinds of systems (including gases, liquids, solids, magnets, and plasmas) and phenomena (such as Brownian motion and black body radiation), and has a remarkably broad application (for instance, in the theory of neural networks[35]).

If the ontology-first approach were assumed, it would be natural to think that the primary task of statistical mechanics is *just* to provide a microphysical foundation for thermodynamics. For, since there are microscopic constituents of thermodynamic systems (that is, there is an ontological relation of reduction), there should be a theory of those constituents that can explain, justify, and in principle make predictions about thermodynamic systems. Such a theory is physically significant *because* it is about the right ontology (i.e., the microscopic constituents of thermodynamic systems) and provides a microphysical foundation for thermodynamics. Any theory that does not

[34] Thanks to Valia Allori for making a suggestion along this line.

[35] For example, see Bahri, Kadmon, Pennington, Schoenholz, Sohl-Dickstein, and Ganguli (2020).

do so would *not* be physically significant and can only be seen as mathematical or instrumental. If the collection of techniques in Gibbsian statistical mechanics is not essentially about the microscopic constituents of thermodynamic systems and their behaviors, one would question if those techniques carried any physical significance at all. The ontology-first approach, accordingly, does not support a broader understanding of statistical mechanics, such as provided by the Gibbsian framework, that goes beyond providing a microphysical foundation for thermodynamics.

In contrast, if the theory-first approach is assumed, a theory establishes its status as a *physical* theory via its usefulness and efficacy. Gibbsian statistical mechanics can thus stand as a successful physical theory on its own, and its physical significance is justified via its usefulness and efficacy (along with providing a microphysical foundation of thermodynamics). Consequently, its broad scope would not be restricted to merely providing a microphysical foundation for thermodynamics.

6 Concluding Remarks

In this paper, I proposed a distinction between the ontology-first and the theory-first approach to reduction. I demonstrated the significance of this distinction by explaining how it plays a role in discussions of Boltzmannian and Gibbsian statistical mechanics: the disagreements between these two frameworks essentially arise from the disagreement between the ontology-first and the theory-first approach. With this in mind, I will discuss in future work whether the ontology-first or the theory-first approach is the *correct* approach to reduction, in the context of reducing thermodynamics to statistical mechanics. Such a discussion can determine, or at least help us gain more insights on, whether the Boltzmannian or the Gibbsian framework is the right framework for statistical mechanics.

The significance of this distinction is not limited to statistical mechanics or physics. Outside statistical mechanics, the distinction may help evaluate whether thermodynamics should be understood as a traditional dynamical theory or as a control theory (Wallace 2014). (This issue is related to but different from the debate discussed in this paper which is on what the right framework of statistical mechanics is.) The distinction can also shed light on discussions on the interpretations of quantum mechanics. Arguably the justification of Bohmian mechanics assumes the ontology-first approach (as I briefly touched on in Section 1.3), whereas the Everettian interpretation needs to assume the theory-first approach. Outside physics, the distinction could potentially be useful for understanding particular cases of reduction in biology, chemistry, philosophy of mind, and so on. Moreover, calling attention to this distinction may help clarify debates surrounding reductionism in general.

Acknowledgements

I would like to thank David Albert, Robert Batterman, Kevin Davey, Samuel Fletcher, Jeffrey Russell, Caitlin Mace, and Wayne Myrvold for reading earlier drafts and for valuable comments and discussions. I am especially grateful to Siddharth Muthukrishnan and David Wallace.

References

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

Allori, V. (2013). Primitive ontology and the structure of fundamental physical theories. In A. Ney and D. Z. Albert (Eds.), *The Wave Function: Essays in the Metaphysics of Quantum Mechanics*, pp. 59–76. Oxford University Press.

Allori, V. (2015). Primitive ontology in a nutshell. *International Journal of Quantum Foundations 1*(3), 107–122.

Allori, V. and N. Zanghi (2004). What is bohmian mechanics. *International Journal of Theoretical Physics 43*, 1743–1755.

Bahri, Y., J. Kadmon, J. Pennington, S. S. Schoenholz, J. Sohl-Dickstein, and S. Ganguli (2020). Statistical mechanics of deep learning. *Annual Review of Condensed Matter Physics 11*(1), 501–528.

Bickle, J. (1996). New wave psychophysical reductionism and the methodological caveats. *Philosophy and Phenomenological Research 56*(1), 57–78.

Boltzmann, L. (1872). Further studies on the thermal equilibrium of gas molecules (weitere studien über das wörmegleichgewicht unter gasmolekülen). *The Kinetic Theory of Gases*, 262–349 (2003).

Boltzmann, L. (1895). On certain questions of the theory of gases. *Nature 51*(1322), 413–415.

Boltzmann, L. (1896). Reply to Zermelo's remarks on the theory of heat ("entgegnung auf die wärmetheoretischen betrachtungen des hrn. e. zermelo"). *The Kinetic Theory of Gases*, 392–402 (2003, Jul).

Boltzmann, L. (1897). On Zermelo's paper "on the mechanical explanation of irreversible processes" (zu hrn. zermelo's abhandlung über die mechanische erklärung irreversibler vorgange). *The Kinetic Theory of Gases*, 412–419 (2003).

Callender, C. (1999). Reducing thermodynamics to statistical mechanics: The case of entropy. *The Journal of Philosophy 96*(7), 348–373.

Callender, C. (2001). Taking thermodynamics too seriously. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 32*(4), 539–553.

Cartwright, N. (1983). *How the Laws of Physics Lie*. Oxford University Press.

Dennett, D. C. (1991). Real patterns. *Journal of Philosophy 88*, 27–51.

Dizadji-Bahmani, F., R. Frigg, and S. Hartmann (2010). Who's afraid of Nagelian reduction? *Erkenntnis 73*(3), 393–412.

Dürr, D., S. Goldstein, and N. Zanghì (1992). Quantum equilibrium and the origin of absolute uncertainty. *Journal of Statistical Physics 67*(5-6), 843–907.

Fodor, J. A. (1974). Special sciences, or disunity of science as a working hypothesis. *Synthese 28*(2), 97–115.

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. Ashgate.

Frigg, R. and C. Hoefer (2015). The best Humean system for statistical mechanics. *Erkenntnis 80*, 551–574.

Frigg, R. and C. Werndl (2019). Statistical mechanics: A tale of two theories. *The Monist 102*(4), 424–438.

Frigg, R. and C. Werndl (2020). Can somebody please say what Gibbsian statistical mechanics says? *The British Journal for the Philosophy of Science*, 1–27.

Goldstein, S. (2001). Boltzmann's approach to statistical mechanics. In J. B. et al (Ed.), *Chance in Physics: Foundations and Perspectives*. Springer.

Goldstein, S. (2012). Typicality and notions of probability in physics. In Y. Ben-Menahem and M. Hemmo (Eds.), *Probability in Physics*, pp. 59–71. Springer.

Goldstein, S. (2019). Individualist and Ensemblist Approaches to the Foundations of Statistical Mechanics. *The Monist 102*(4), 439–457.

Goldstein, S., J. L. Lebowitz, R. Tumulka, and N. Zanghi (2019). Gibbs and Boltzmann entropy in classical and quantum mechanics. https://arxiv.org/abs/1903.11870.

List, C. (2019). Levels: Descriptive, explanatory, and ontological. *Noûs 53*(4), 852–883.

Malament, D. B. and S. L. Zabell (1980). Why Gibbs phase averages work—the role of ergodic theory. *Philosophy of Science 47*(3), 339–349.

Maudlin, T. (1995). Review [of Sklar, time and chance and Sklar, philosophy of physics]. *The British Journal for the Philosophy of Science 46*(1), 145–149.

Maudlin, T. (2007). Completeness, supervenience and ontology. *Journal of Physics A: Mathematical and Theoretical 40*, 3151–3171.

Maudlin, T. (2010). Can the world be only wavefunction? In S. Saunders, J. Barrett, A. Kent, and D. Wallace (Eds.), *Many Worlds?: Everett, Quantum Theory and Reality*. Oxford University Press.

Maudlin, T. (2016). Local beables and the foundations of physics. In M. Bell and S. Gao (Eds.), *Quantum Nonlocality and Reality: 50 Years of Bell's Theorem*, pp. 317–330. Cambridge University Press.

Maudlin, T. (2019). *Philosophy of Physics: Quantum Theory*. Princeton University Press.

McCoy, C. D. (2020a). An alternative interpretation of statistical mechanics. *Erkenntnis 85*(1), 1–21.

McCoy, C. D. (2020b). Interpretive analogies between quantum and statistical mechanics. *European Journal for Philosophy of Science 10*(1), 9.

McIntyre, L. (2007). Emergence and reduction in chemistry: Ontological or epistemological concepts? *Synthese 155*(3), 337–343.

Moulines, C. U. (1984). Ontological reduction in the natural sciences (1). In W. B. A. P.-J. Schmidt (Ed.), *Reduction in Science: Structure, Examples, Philosophical Problems*, pp. 51–70. D. Reidel Publishing Company.

Nagel, E. (1961). *The Structure of Science: Problems in the Logic of Scientific Explanation*. Harcourt, Brace and World.

Ney, A. (2013). Ontological reduction and the wave function ontology. In A. Ney and D. Z. Albert (Eds.), *The Wave Function: Essays on the Metaphysics of Quantum Mechanics*, pp. 168–183. Oxford University Press.

Oppenheim, P. and H. Putnam (1958). Unity of science as a working hypothesis. *Minnesota Studies in the Philosophy of Science 2*, 3–36.

Poidevin, R. L. (2005). Missing elements and missing premises: A combinatorial argument for the ontological reduction of chemistry. *British Journal for the Philosophy of Science 56*(1), 117–134.

Robertson, K. (2018). Asymmetry, abstraction, and autonomy: Justifying coarse-graining in statistical mechanics. *The British Journal for the Philosophy of Science 71*(2), 547–579.

Rohrlich, F. (1988). Pluralistic ontology and theory reduction in the physical sciences. *British Journal for the Philosophy of Science 39*(3), 295–312.

Salmon, W. C. (1984). Scientific explanation: Three basic conceptions. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association 2*, 293–305.

Sarkar, S. (1992). Models of reduction and categories of reductionism. *Synthese 91*(3), 167–94.

Schaffer, J. (2008). Causation and laws of nature: Reductionism. In T. Sider, J. Hawthorne, and D. W. Zimmerman (Eds.), *Contemporary Debates in Metaphysics*, pp. 82–107. Blackwell.

Schaffner, K. F. (2012). Ernest Nagel and reduction. *Journal of Philosophy 109*(8-9), 534–565.

Sklar, L. (1993). *Physics and Chance: Philosophical Issues in the Foundations of Statistical Mechanics*. Cambridge University Press.

Smart, J. (1959). Sensations and brain processes. *Philosophical Review 68*, 141–156.

Uffink, J. (2007). Compendium of the foundations of classical statistical physics. In J. Butterfield and J. Earman (Eds.), *Philosophy of Physics*, pp. 923–1047. Amsterdam: North Holland.

van Fraassen, B. C. (1980). *The Scientific Image*. Oxford University Press.

van Gulick, R. (2001). Reduction, emergence and other recent options on the mind/body problem: A philosophic overview. *Journal of Consciousness Studies 8*(9-10), 1–34.

van Riel, R. (2014). *The Concept of Reduction*. Springer International Publishing.

van Riel, R. and R. van Gulick (2019). Scientific reduction. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2019 ed.). Metaphysics Research Lab, Stanford University.

Wallace, D. (2014). Thermodynamics as control theory. *Entropy 16*, 699–725.

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

Wallace, D. (2020). The necessity of Gibbsian statistical mechanics. In V. Allori (Ed.), *Statistical Mechanics and Scientific Explanation: Determinism, Indeterminism and Laws of Nature*. World Scientific Publishing Co.

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

Zwanzig, R. (1960). Ensemble method in the theory of irreversibility. *The Journal of Chemical Physics 33*(5), 1338–1341.

Zwanzig, R. (1966). Statistical mechanics of irreversibility. In P. Meijer (Ed.), *Quantum Statistical Mechanics*. New York: Gordon and Breach.