The London School of Economics and
Political Science

Identity and Indistinguishability in Thermal Physics

James Wills

A thesis submitted to the Department of
Philosophy of the London School of Economics
and Political Science for the degree of Doctor of
Philosophy, London, May 2022.

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it). The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made. This thesis may not be reproduced without my prior written consent. I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party. I declare that my thesis consists of 49497 words.

Abstract

This thesis proposes mathematically precise analyses of the concepts of identity and indistinguishability and explores their physical consequences in thermodynamics and statistical mechanics. I begin by exploring the philosophical consequences of the geometric formulation of thermodynamics, well-known to many mathematicians. Based on this, I offer novel accounts of what it means to be a thermodynamic system and what it means to be a composite system. I then use these mathematical tools to offer new and precise definitions of 'mixture' and 'identity' in thermodynamics. These analyses allow me to propose a novel resolution of Gibbs' paradox. Finally, I offer a new definition of indistinguishability in statistical mechanics with a view to offering a new resolution of Gibbs' paradox in statistical mechanics (the $N!$ problem). My analysis highlights the importance of observables in the foundations of statistical theories.

Contents

Chapter 1

Introduction

Problems once deemed to be the domain of physics have become philosophical. Part of the reason for this is that physics and philosophy are now separate academic disciplines where there was only one discipline before, sometimes known as *natural philosophy*. This encompassed questions and debates, some of which would now be considered strictly the realm of physics or philosophy. Of course, this divergence in the disciplines should come as no surprise; as our knowledge about the natural world and our place in it grows, so does our ignorance, and branching academic disciplines form to encompass these new questions. Looking back from our temporal vantage point, topics we would now consider strictly the domain of philosophy were studied by people who would now be considered physicists. We do not have to look too deeply in the history of science to see that this is true. Newton is an easy example: his *Principia* is now seen as the beginning of the physics of classical mechanics and gravitation, but it also contains debates and questions which are now uncontroversially seen as philosophical: the debate concerning the nature of space and time (the relationism-substantivalism debate) is a prime example. Einstein is yet another example: his work on relativity theory gave rise to an enormous philosophical literature and many passages of his work may be considered as touching on philosophical topics. And yet, in writing these passages, Einstein may well have thought he was doing physics.

Thus, some problems once deemed physical have become philosophical. This thesis concerns such a problem, one which arose in the context of the science of thermodynamics, but gradually caught the attention of philosophers. It concerns the apparently niche puzzle of calculating the entropy increase when gases are mixed. The standard (and basically correct) line is that there is a non-zero entropy increase when the gases are different but no entropy change when the gases are the same. The philosophical puzzle enters when seek further clarity on how best to represent mixtures, what the concepts 'different' and 'the same' should mean, and how they should be precisely defined. This cluster of puzzles has come to be known as *Gibbs' paradox*.

While Gibbs himself did not describe the puzzle he posed as a paradox, the gas mixing scenario which gave rise to subsequent discussion is first described in his monumental 1878 work on the foundations of thermodynamics, *On the Equilibrium of Heterogeneous Substances*. He later revisited gas mixing in the very last paragraph of his equally monumental 1902 work on the foundations of statistical mechanics, *Elementary Principles of Statistical Mechanics*.

It was first dubbed a 'paradox' in a paper by O. Wiedeburg entitled "The Gibbs paradox". Although 'Gibbs' paradox' was originally intended to refer only to the thermodynamic puzzle, it came to refer, thanks to Gibbs' mention of gas mixing in the final paragraph of Gibbs (1902), to a totally different puzzle in statistical mechanics also known as the '$N!$ problem'. This, in brief, concerns the justification of the factor $N!$ in the state of a statistical mechanical system consisting of N 'indistinguishable' particles. It is argued that this factor is important in ensuring an extensive entropy function. Gibbs himself did not appear to see a problem here, although his prose when explaining why the $N!$ factor appeared was not always clear in critical places of his exposition. This no doubt contributed to the confusion surrounding the '$N!$ problem' version of the Gibbs paradox.

It is unclear who first used 'Gibbs' paradox' to refer to the $N!$ problem. In any case, the use of the term to refer to the $N!$ problem soon outweighed the thermodynamic use of the term given to it by Wiedeburg; throughout the twentieth century, the real puzzle was felt to be the justification of the $N!$ factor, partly because it was felt to be the more interesting version of the paradox and partly because it was felt that the thermodynamic version of the paradox was solved.

Throughout the twentieth century, physicists and textbook writers felt the need to comment on the $N!$ problem in connection with indistinguishable particles, some letting their emotions get the better of them. For example, Huang (1963, p. 141) called the problem a "disastrous result" and Schrödinger (1989, p. 61) dramatically wrote:

> "After a railway accident, or a fire, or a similar disaster, the authorities are always anxious to answer the question: How could it have happened?" Schrödinger, 1989, p. 61

Such a great physicist lamenting the existence and persistence of this problem should provide reason enough to think about it. The urgency of the problem continues to be felt to this day:

> "Indeed, [the $N!$ problem] is so serious that classical macroscopic ther-

> modynamics fails at this point since it does not recognise the concepts of distinguishability and indistinguishability!" (Glazer and Wark, 2001, p. 59)

And so the Gibbs paradox persists through time, occupying short sections in almost every textbook. Unfortunately, this means that there are almost as many statements and solutions to the problem as there are authors, making a complete dissection of the problem almost impossible. However, one historical development of the paradox is worthy of note.

Insofar as they were problems at all in Gibbs' eyes, the thermodynamic and statistical mechanical versions of the paradox were originally stated in classical physics: quantum theory was barely on the horizon when Gibbs published his final work on statistical mechanics at the turn of the twentieth century. Despite this, many authors take quantum theory to be essential in solving the $N!$ problem, on the basis that only in quantum theory can particles be, in some sense, 'indistinguishable'. This view is nicely represented in this passage from Dugdale (1966):

> "According to the point of view of quantum mechanics, molecules or atoms cannot even in principle be continuously observed so that they cannot maintain a continuous identity. Consequently molecules or atoms of the same kind in an assembly are to be thought of as fundamentally indistinguishable." (Dugdale, 1966, p. 92)

In this passage, Dugdale leaves unsaid the exact feature of quantum theory which characterises the indistinguishability. Indeed, even on this there is no consensus. Huang (1963) thinks that the $N!$ problem cannot be solved by classical mechanics alone and states that the relevant feature of quantum theory which solves the puzzle is the Symmetrisation Postulate:

> "It is not possible to understand classically why we must divide [. . .] by $N!$ to obtain the correct counting of states. The reason is inherently quantum mechanical. Quantum mechanically, atoms are inherently indistinguishable in the following sense: A state of the gas is described by an N-particle wave function, which is either symmetric or antisymmetric with respect to the interchange of any two particles. A permutation of the particles can at most change the wave function by sign, and it does not produce a new state of the system. Hence we should divide [. . .] by $N!$. This rule of counting is known as the "correct Boltzmann counting". It is something that we must append to classical mechanics in order to get right answers." (Huang, 1963, p. 141)

Ter Haar also clearly indicates that the solution to the $N!$ problem is to be found in quantum mechanics:

> "The resolution of this paradox is [. . .] another example of the successes of quantum mechanics in getting rid of some unsolved problems of classical physics" (ter Haar, 1954, p. 146)

However, in contrast to Huang he argues that the indistinguishability of the particles is to be cashed out in terms of Heisenberg's Uncertainty Principle:

> "So long as we are dealing with classical systems, the $N!$ microsituations are in fact different and can be distinguished from one another. In quantum mechanical systems, however, we can no longer distinguish

> them. This is certainly due to the fact that we cannot follow the atoms so closely along their orbits that we can, for instance, know which atom is which after a collision. This again is a consequence of Heisenberg's famous relations." (ter Haar, 1954, p. 72)

Thus we see at least two eminent physicists declaring that the $N!$ problem, despite being originally stated in the context of classical physics, requires an inherently quantum mechanical solution. One may speculate that they consider it to be a similar type of problem to others which precipitated the development of quantum mechanics, such as the ultraviolet catastrophe or the problem of specific heats and degrees of freedom at low temperatures[1], those which Kuhn, in his account of scientific revolutions, would have called 'anomalies'.

This view is not implausible. However, subsequent development of the view became so when Gibbs was attributed almost clairvoyant capabilities:

> "The new factor $1/N!$ has arisen from our consideration of the symmetry of wave functions. In turn, the quantum mechanical postulate on symmetry of wave functions has its origin in the experimental indistinguishability of identical particles. Gibbs has the intuitive foresight, before the advent of quantum mechanics, to insert this factor $1/N!$." (Hill, 1956)

This passage from Terrell Hill's textbook suggests the $N!$ factor to be some kind of 'correction' which had to be put in 'by hand' and that Gibbs inserted it, knowing that it was necessary, but without really knowing why, before quantum theory came

---

[1] This particular issue is discussed in Huang (1963, p. 138) and I speculate that he may have thought the $N!$ problem to be of a similar kind to it.

along with the final and correct explanation. Similarly, Bernard Lavenda, drawing directly on Gibbs' final chapter of *Elementary Principles*, writes:

> "In Chapter XV of Elementary Principles in Statistical Mechanics, Gibbs asked the following question: "If two phases differ only in that certain entirely similar particles have changed places with one another, are they to be regarded as identical or different phases?" His answer was that "if the particles are regarded as indistinguishable, it seems in accordance with the spirit of the statistical method to regard the phases as identical." With these words, Gibbs not only introduced the concept of indistinguishability of identical, nonlocalizable particles but also foresaw the advent of quantum mechanics." (Lavenda, 1991, p. 223)

I will willingly grant that Gibbs was one of the most astute and intelligent physicists in history, but I am less willing to place the same interpretation on Gibbs' words as Lavenda.

The effect of the history of the Gibbs paradox on the physics and philosophy literature has been two-fold. (1) An overwhelming focus on the $N!$ problem as the more 'interesting' version of the paradox. (2) An overwhelming focus on quantum theory for solving the $N!$ problem. One way of viewing this thesis is as some way of redressing this balance. The first three chapters concern the theory of classical thermodynamics and the thermodynamic version of Gibbs' paradox. The final chapter concerns classical statistical mechanics and the $N!$ problem. In short, there are a lot more interesting ideas in thermodynamics and classical physics than the history of the Gibbs paradox suggests. In this thesis, we shall see that reflecting on Gibbs' paradox, while it might seem an extremely niche problem to consider, forces us to reflect profoundly on the theory in which the problem is stated and set; as well as learning about Gibbs' paradox, we will learn about the foundations of classical thermodynamics and statistical mechanics.

Characteristic of this type of problem on the cusp of physics and philosophy is the sense that neither discipline's methodologies is able to give a fully satisfactory account of it: the problem requires physical insight, mathematical precision and philosophical clarity. This thesis will marshal methodologies from these fields to answer the following questions:

- What is the mathematical structure of thermodynamics?
- What is the best representation of mixtures as thermodynamic systems?
- How do we give the concepts of identity and indistinguishability in a clear and precise meaning in thermodynamics and classical statistical mechanics?

While there has been some attention given to the philosophical and mathematical foundations of thermodynamics and to the meaning of identity and indistinguishability, I hope to contribute to this literature by deploying existing, but philosophically little-known, mathematical tools to offer new answers and insights to these philosophical questions.

1.1 The mathematical structure of thermodynamics

Vladimir Arnold once wrote:

> "Every mathematician knows that it is impossible to understand any elementary course in thermodynamics." (Arnold, 1990, p. 163)

This may come as a surprise to many physicists who learned thermodynamics in their early undergraduate studies and found very little trouble getting to grips with the simple differential calculus required to solve the problems. However, if one thinks about anything too carefully one is bound to come up against problems. And this is exactly what happens when we ask questions such as: if the mathematical structure of a classical mechanical system is a symplectic manifold and the mathematical structure of a quantum system is a Hilbert space with an algebra of observables, what is the mathematical structure of a thermodynamic system? We quickly find that our undergraduate learnings of nineteenth century thermodynamics of Clausius, Kelvin, Planck et al. does not equip us with the understanding we require to answer such questions.

The Gibbs paradox was originally seen as a problem set and stated in classical thermodynamics. In order to fully understand the problem and its solution, we have to understand the theory in which it is set. Chapter 2 is all about how we may not understand classical thermodynamics as well as our plethora of textbooks may lead us to believe. Arnold, however, quickly absolves us of all guilt and foolishness we may feel for not understanding thermodynamics when we thought we did, for, directly after the quotation above, he goes on to write:

> "The reason is that the thermodynamics is based—as Gibbs has explicitly proclaimed—on a rather complicated mathematical theory, on the contact geometry." Arnold, 1990, p. 163

Arnold goes on to introduce the contact geometric formulation of thermodynamics, arguing that it captures Gibbs' mathematical treatment of the theory, despite the abstract geometric apparatus not being available to him at the time.

While this formulation is not new to mathematicians, it is new to philosophers, so I spend time discussing some potential philosophical and foundational consequences of the theory. The first significant consequence is the identification of a fundamental distinction, explained in Burke (1985), between 'unobserved' or 'hidden' degrees of freedom and 'observed' degrees of freedom. This distinction grounds the difference between the entropy and the other extensive variables and therefore offers a novel and clear interpretation of the thermodynamic entropy. The position advocated here is similar to, but nevertheless importantly different from, claims made in the philosophical literature discussing the anthropocentricity or not of thermodynamics by, for example, Myrvold (2020).

The second significant consequence is the identification of the fundamental relation and its associated first-order homogeneity as a crucial part of defining what it means to be a thermodynamic system. This homogeneity property plays a key role later on in Chapter 4 when I introduce my definition of identity of thermodynamic systems.

Other foundational consequences include clarification of the distinction between intensive and extensive variables, mathematically precise definitions of 'thermodynamic system' and 'composite system' and a clear definition of thermodynamic equilibrium based on the convexity property of the energy function (or, equivalently, the concavity property of the entropy function).

1.2 How to represent mixtures in thermodynamics

Gibbs' paradox is all about mixtures and how the entropy changes when gases mix. My aim is to show that the geometric formulation of thermodynamics explained in the previous chapter can offer new insights into Gibbs' paradox, so the first step in this argument has to be establishing how mixtures and the mixing process are modelled in the formalism.

Using this new formulation, I argue that there exist two competing representations of mixtures in the literature: the 'standard' representation, widely used in presentations of Gibbs' paradox in thermodynamics, and the 'Gibbs' representation, a model used explicitly by Gibbs in his 1878 work. It is striking that Gibbs' representation, written in the very same work where the original form of the Gibbs paradox appeared, is not also the most widely used representation. The important difference between them is that the standard representation treats the 'partial pressures' of the gases as real thermodynamic variables whereas the Gibbs representation has only one pressure variable for the mixture.

To adjudicate between them, I employ two plausible philosophical principles about representation inspired by discussions in the foundations of spacetime theories. I show that, in order to adhere to them both, we must adopt the Gibbs representation over the standard representation. Making this change means we have to rethink how to model the mixing process. In orthodox thermodynamics, mixing is modelled with semipermeable membranes which make use of the partial pressures of each component of the mixture to model the mixing. Since the Gibbs representation of mixture has no partial pressures, we can no longer use this model.

To plug this gap, I adopt the definitions of composite system and thermodynamic equilibrium developed in Chapter 2 to model mixtures as composite systems, and to model mixing as equilibration of subsystems of the composite.

Adopting Gibbs' representation of mixtures has further consequences for thermodynamics: with the elimination of partial pressures from the definition of mixture, theorems of thermodynamics concerning mixtures, such as Dalton's Law about the pressures of mixtures and Gibbs' Theorem concerning the entropy of mixtures, have to be restated.

1.3 How to represent identity in thermodynamics

The main purpose of Chapter 4 is to introduce a new definition of thermodynamic identity and to use it to solve Gibbs' paradox in thermodynamics. Problems concerning 'identity', 'indistinguishability' and other similar terms such as 'indiscernible' or 'individual' have a long history in analytic philosophy, arguably going back to Leibniz and his Principle of the Identity of Indiscernibles. Since this time, the philosophy of identity has been rife with debate over what exactly these terms mean.

Use of these terms should come with a metaphysical health hazard warning. It is unfortunate that the very same word carries widely different meanings depending on who one reads or, worse, it is simply assumed that the meaning of the word is obvious and is not defined with sufficient precision or without regard to the way other users define the term. This, in itself, is philosophical reason enough to get involved in this literature. However, in this work, I am not concerned with delineating the meanings of these words for fear of getting myself stuck in a metaphysical and linguistic quagmire. I am only concerned with these concepts insofar as they are useful to us as physicists. I shall therefore simply take care to give precise and physically relevant definitions of ‘identity’ and ‘indistinguishability’. It may well be the case that what I mean by ‘indistinguishability’, someone else calls ‘identity’ or what I mean by ‘indistinguishability’ is what someone else calls ‘indiscernibility’. Ultimately, it does not matter how we use certain words, so long as we take sufficient care to say exactly what we mean by them.

I begin by stating Gibbs’ paradox in thermodynamics and then I describe and critically analyse an extremely popular response by van Kampen which has guided most of the philosophical discussion on this version of the paradox and is used as a prime example of the ‘operational’ or ‘anthropocentric’ nature of thermodynamics[2]. My main contribution to the literature on this topic is guided by the following methodological principle: the concepts used by a theory should be expressible in the theory’s formalism and should not appeal to theoretic elements outside the theory’s domain. For example a proposition such as “gases A and B are identical” or “gases A and B are different” are treated as thermodynamic propositions and assumptions in the Gibbs paradox setup. If we want them to be genuine thermodynamic propositions, we should strive to express the identity relation in purely thermodynamic terms. This means not appealing to non-thermodynamic concepts like, for example, proton number.

I argue that the main difficulty with van Kampen’s approach is that the criterion of identity he adopts is extra-theoretic, in the sense that it appeals to concepts

---

[2] See, for example, Myrvold (2020), Saunders (2013), and Dieks (2018)

(such as semipermeable membranes) which thermodynamics does not have the capacity to represent.

I go on to offer a thermodynamic definition of identity, based on the homogeneity of thermodynamic systems, which is similar in spirit to other definitions of identity given in more modern discussions in formal philosophy (see Halvorson (2019)). In short, two thermodynamic systems are identical when they each scale up to a third thermodynamic system which may be viewed as the composite of the two. I show how this makes sense of Gibbs' paradox by providing clear and precise definitions of difference and identity.

1.4 How to represent indistinguishability in classical statistical mechanics

In Chapter 5, I tackle the statistical mechanical version of the Gibbs paradox, or the $N!$ problem. I propose a purely classical, non-quantum, solution to the problem, in line with the thinking of other philosophers of physics and in contrast to many physicists. I base my solution on two novel foundational points.

The first concerns the general interpretation of statistical theories. I argue that the sample spaces on which probability measures are defined should be constrained by what is observable in order for our measurements of probabilities as frequencies to make sense. The second point supplies a definition of classical particle indistinguishability in terms of classical observables (which are functions $f : M \to \mathbb{R}$ where M is the phase space). Taken together, I show that we can derive the factor $N!$ through rigorous mathematical argument, contrary to the suggestions of authors quoted above that the $N!$ factor is some kind of 'correction' which has to be inserted by hand. In the same way that Chapter 4 argues the thermodynamic identity of gases must be expressed purely in the formalism of thermodynamics without appealing to extra-theoretic concepts, so in Chapter 5 the indistinguishability of particles must be expressed purely in the formalism of classical statistical mechanics. In each case, this approach allows us to formulate mathematical propositions about the identity and indistinguishability of gases which may be used to derive theorems about the entropy of mixing and about the factor $N!$ respectively.

The framework I develop on the basis of these two points helps us understand and compare similar solutions to the $N!$ problem against the same formal background and also provides a natural bridge to thinking about indistinguishable quantum particles. I anticipate that the framework developed in this chapter will be of great use in thinking about indistinguishability in quantum theory and the senses in which it differs from classical mechanics.

Chapter 2

Thermodynamics and Contact Geometry

2.1 What is Thermodynamics?

This question is difficult to answer. But the difficultly faced in answering it is well-rewarded in the form of a mathematically precise and philosophically deeper understanding of thermodynamics. As far as standard answers to this question go, thermodynamics lags far behind other mathematised physical theories in mathematical and conceptual precision. To illustrate, let us consider the answers to the same question asked of other physical theories. What is classical Hamiltonian mechanics? It is the study of classical Hamiltonian systems, represented by the triple (M, h, ω) where M is a $2n$-dimensional manifold which can be given local coordinates $q_1, ..., q_n, p_1, ..., p_n$ and n is the number of degrees of freedom of the system, h is the Hamiltonian function defined on M, and ω is the symplectic form. What is quantum mechanics? It is the study of quantum systems represented by the pair $(\mathcal{H}, \rho)$ where $\mathcal{H}$ is a complex vector space with inner product, the Hilbert space, and ρ is a density matrix defined on $\mathcal{H}$. What is general relativity? It is the study of spacetimes represented by the pair (M, g_{ab}), where M is a manifold and g_{ab} is a metric on M, and the dynamical interaction between the metric and matter fields given by Einstein's field equation.

Now, what is thermodynamics? This does not have such a concise answer. A survey of answers to this question from various commentators indicates that there are three main strands. These are:

1. The science of the relationship between heat and work.
2. The science of the macroscopic consequences of the very large number of microscopic degrees of freedom.
3. The science of the relationship between the observed degrees of freedom and the unobserved degrees of freedom.

These, I think, are placed roughly in order of familiarity; the first answer is what the majority of physicists learn on being introduced to thermodynamics for the first time, while the second and third are less well known. In the following sections, I will expand on these three characterisations. My exposition of the literature and history will necessarily be selective and therefore incomplete; I focus only on those aspects which give us a clue as to each view's basic philosophical commitments, with a view to motivating and defending geometric thermodynamics as my preferred formulation.

In Section 2.1.1, I will trace some important episodes in the history of the first characterisation and in Section 2.1.2, I discuss the second and third characterisations. We will learn that the second is linked to the plausible assumption that statistical mechanics is the correct reductive base for thermodynamics. I will argue that the third characterisation provides the most general and unifying answer to our question.

2.1.1 The science of heat and work

The main inspiration for this characterisation of thermodynamics is historical: thermodynamics was born out of investigations into how heat can be converted into work and vice versa. I will not attempt to give a full history of thermal physics[1] but I shall sketch the history enough to illustrate the sense in which the distinction between heat and work came to be seen as fundamental to thermodynamics.

At around the time thermodynamics was beginning to be developed, there was a great deal of industrial interest in using powerful engines to move goods and automate processes. The steam engine was new technology and engineers were working hard to adapt it to a variety of tasks. It is no surprise that scientists became interested in this; industrialists wanted their engines to be as efficient as possible, extracting the maximum amount of useful work (and profit) for the minimum cost of fuel, engineers helped the industrialists by developing and honing the new technology to achieve this and the scientists helped the engineers by developing the theoretical underpinnings of the new technology.

In fact, the first working engines which converted heat from the burning of fuel into useful work were in use long before the theoretical workings of those engines were properly understood. Sadi Carnot's famous book *Reflections on the Motive Power of Fire* was inspired by the practical consideration of understanding the scientific principles behind the steam engines prevalent in the industrial world at

---

[1] There are many excellent historical texts (e.g. Cardwell (1971) and Truesdell (1980)).

the time. He noticed that these principles were not properly understood, and that any improvement in the science of steam engines was brought about more by trial and error than methodical investigation:

> "Notwithstanding the work of all kinds done by steam-engines, notwithstanding the satisfactory condition to which they have been brought to-day, their theory is very little understood, and the attempts to improve them are still directed almost by chance." (Carnot, 1872/1890, p. 42)

He therefore set himself the task of understanding these principles, ultimately with the hope of engineering better engines, but with the immediate intellectual satisfaction of understanding the principles that govern them. One of the most important contributions he made in his book was to abstract away many of technical and engineering subtleties of steam engines and consider the production of work from heat from an extremely general point of view:

> "In order to consider in the most general way the principle of the production of motion by heat, it must be considered independently of any mechanism or any particular agent. It is necessary to establish principles applicable not only to steam-engines but to all imaginable heat-engines, whatever the working substance and whatever the method by which it is operated." (Carnot, 1872/1890, pp. 43–44)

Carnot thus coined the term *heat engine*: simply a device which extracts work from heat, without any consideration of what it is made of or how it does it. This allowed him to see past ultimately irrelevant details and hone in on the basic feature of steam engines which allowed them to extract work from heat. This feature, which Carnot was first to isolate, was this: work can only be extracted when there is a temperature gradient. In other words, there must be both a hot and a cold body for heat to pass between in order for work to be extracted from heat. This observation was common to all heat engines at the time and was recognised by Carnot as being the essential feature which explained why they work.

Granted that this fact explains why the engines work, but why is it true? What explains this fact? What is the nature of heat such that it can produce work when it passes from a hot body to a cold body? In order to answer these questions, we need to adopt a theory of heat: an account of the nature of heat and how it behaves. Before the science of thermodynamics settled into the form we currently understand and are taught today, there were two theories of thermodynamics based on two distinct views on heat. One was the *caloric theory*, which postulated that heat was a kind of conserved, weightless, invisible fluid which flowed from the hotter substance, making it cooler, to another cooler substance, making it hotter. This flow was understood to give rise to useful mechanical work (in the form of the motion of a piston, for example), in much the same way as water falling from a height imparts motion to a water mill. This theory had much success explaining common heat phenomena, such as the equilibration of temperatures of two bodies in thermal contact and the ability to extract work from temperature differences. The other was the *kinetic theory*, which postulated that heat was the motion of the particles which made up a substance. This theory was also able to account for many of the same phenomena as the caloric theory, including being able to explain how heat is produced by friction.

Carnot adopted the caloric theory throughout his book and was able to explain how work was extracted from heat passing from hot to cold as follows:

> "The production of motive power is then due in steam-engines not to an actual consumption of caloric, but to its transportation from a warm body to a cold body, that is, to its re-establishment of equilibrium—an equilibrium considered as destroyed by any cause whatever, by chemical action such as combustion, or by any other. We shall see shortly that this principle is applicable to any machine set in motion by heat.
>
> According to this principle, the production of heat alone is not sufficient to give birth to the impelling power: it is necessary that there should also be cold; without it, the heat would be useless." (Carnot, 1872/1890, p. 46)

It is important to note the following consequence of adopting the caloric theory: being a conserved material substance, caloric is not consumed, it merely moves from hot to cold. As Carnot notes, this explains why both a hot and a cold body is necessary to extract 'motive power'; with only one body, the caloric cannot move anywhere and without the caloric moving, no work can be extracted. It is clear from this discussion of the caloric theory that a fundamental distinction between heat and work is set firmly at the foundations of caloric thermodynamics: heat as a material substance and motive power, or work, as a form of energy.

As it happened, the kinetic theory of heat eventually came to replace the caloric theory. How this happened is a fascinating story but is not my main concern.[2] I am, however, concerned with understanding how the distinction between heat

---

[2] Excellent accounts may be found in Cardwell (1971) and in Psillos (1999, pp. 115–130) and Chang (2003) who tell the story from the point of view of the scientific realism debate.

and work came to be seen as fundamental to the kinetic theory and what exactly grounds this distinction.

It is tempting to think of the transition from the caloric theory to the kinetic theory as a dramatic paradigm shift, with the caloric theory dismissed unceremoniously and the kinetic theory held high on the shoulders of the scientific community as the new kid on the block able to solve all the problems left behind by the caloric theory. The reality of the transition is rather different. While it is true that the caloric theory was for some time the favoured account of heat, both the caloric and kinetic theories existed in tandem in the scientific community and each got serious consideration from scientists. The caloric theory was also deemed highly successful in explaining heat phenomena which altogether lessened the appetite for theory change. Nevertheless, observational evidence eventually stacked up against the caloric and for the kinetic theory. The exact chain of events leading to this is a matter for serious historical study, but it is certain that the following key observations pushed scientists to adopting the kinetic theory in favour of the caloric theory:

1. Heat can be generated from friction. Caloricists had an explanation for this: the work done causes the "destruction of the equilibrium of the caloric" (Carnot, 1872/1890, p. 51), reciprocally to the observation that the equilibration of the caloric enables work to be done. However, the kinetic theory gave what transpired to be a 'better' explanation: the work done is converted into the kinetic energy of the particles of the substance.

2. Related to the first, James Joule performed experiments (famously his 'paddle-wheel' experiment) which showed that the heat generated by friction was

always in constant proportion to the work done. This indicated that the work is converted into heat and is thus a kind of energy.

3. The failure to measure the weight of the caloric. If heat were a material substance, so the argument goes, it ought to have some weight. However, experiments performed to find its weight by Count Rumford failed to find any weight ascribed to heat. By contrast, the kinetic theory was easily able to accommodate the fact that a hot body weighs the same as when it is cold.

As Clausius puts it:

> “These circumstances, of which Carnot was also well aware, and the importance of which he expressly admitted, pressingly demand a comparison between heat and work, to be undertaken with reference to the divergent assumption that the production of work is not only due to an alteration in the distribution of heat, but to an actual consumption thereof; and inversely, that by the consumption of work heat may be produced. (Clausius, 1851, p. 2)

Thus the transition to the kinetic theory was made. While it was a step away from the caloric theory, Clausius did not see it as a large one:

> “It is not even requisite to cast the theory of Carnot over-board; a thing difficult to be resolved upon, inasmuch as experience to a certain extent has shown a surprising coincidence therewith. On a nearer view of the case, we find that the new theory is opposed, not to the real fundamental principle of Carnot, but to the addition “no heat is lost;” for it is quite possible that in the production of work both may take

> place at the same time; a certain portion of heat may be consumed, and a further portion transmitted from a warm body to a cold one; and both portions may stand in a certain definite relation to the quantity of work produced. (Clausius, 1851, p. 4)

It is clear from this passage that Clausius did not see the kinetic theory of heat as entailing the rejection of Carnot's work altogether. This is fortunate, he writes, given how successful it was at accommodating the facts. It is not even necessary to reject 'the real fundamental principle of Carnot' by which he means the principle that work can be extract from the flow of heat from hot to cold. This principle is essentially correct; all we have to do is amend the clause 'no heat is lost in the flow' to 'some heat is transferred and the rest is converted to work'. This amendment amounts to a rejection of the caloric theory (where heat is conserved) to the kinetic theory (where heat may be 'consumed', i.e. converted to work). As a result of this transition, Clausius formulated the science of thermodynamics more or less as we are taught it today.

The purpose of viewing the transition from the caloric theory to the kinetic theory from this angle, is to illustrate just how much was preserved in the transition. In particular, the transition required no change in what thermodynamics was fundamentally about; it is still concerned with the relationship between work and heat and how one may be obtained from the other. In other words, the distinction between heat and work remains fundamental under the transition from the caloric to the kinetic theory.

To further reinforce the point, we can investigate the role of the distinction in the first and second laws of thermodynamics as Clausius formulated them, and as are still learned today. Clausius formulated the first law of thermodynamics, which he called The First Main Principle of the Mechanical Theory of Heat, as follows:

> "In all cases where work is produced by heat, a quantity of heat consumed is proportional to the work done; and inversely, by the expenditure of the same amount of work the same quantity of heat may be produced." (Clausius, 1879, p. 23)

It is clear that this proposition relies on a distinction between heat and work. This statement is a special case of the principle of energy conservation and applies to cases where the only forms of energy transfer are as work or heat. In his subsequent mathematical development of the law, he introduces the concept of the *internal energy*. In standard expositions of thermodynamics, the internal energy is introduced by appealing to our knowledge and intuitions about the microscopic physics of the thermodynamic system: it is defined to be the sum of the energy of all the internal, microscopic degrees of freedom of the system. From this perspective, it is clearly by definition a function of state. This method of introduction, however true it is from a microphysical perspective, relies for its definition on microscopic physics, something we should be avoiding if we are attempting to introduce and define thermodynamic quantities without appealing microscopic physics. We can avoid this definition of internal energy by following Clausius in providing a purely thermodynamic characterisation, on which we may subsequently impose a microscopic interpretation. We can see how Clausius' form of the first law implies the existence of a state function on to which we may impose such a microscopic interpretation.

To show this, it is necessary to appreciate the significance of cyclic processes in thermodynamics. These processes take a thermodynamic system through a series of changes of state before returning the system to its initial state. Over the course of these changes, energy in the form of work or heat may enter or leave the system, but the system remains in the same state at which it began. This means that the net work done on the system and the net heat entering the system over the course of the cycle has no effect whatsoever on the system itself. So, if we study only cyclic processes, we do not have to worry about analysing the energy changes of the system itself. If we apply Clausius' first main principle to cyclic processes we get the following proposition: if a cyclic process is performed such that work is done on/by the system, then an equal amount of heat energy left/entered the system. This is expressed mathematically as follows:

$$\oint dW = \oint dQ \tag{2.1.1}$$

from which it follows that

$$\oint dW - dQ = 0. \tag{2.1.2}$$

Since the integral of the differential $dW - dQ$ is zero over the course of a cycle, it is by definition a state function. We may introduce the symbol U for this function and define it thus:

$$dU := dW - dQ \tag{2.1.3}$$

where dW is the work done on the system and dQ is the heat leaving the system.

Thus we arrive at the canonical formulation of the first law of thermodynamics and we can see clearly how it follows from Clausius' First Main Principle. Set up in this way, we can see that Clausius' First Main Principle implies the existence of the state function U.

The distinction between heat and work is also crucial to the formulation of the second law of thermodynamics. Calling it the Second Main Principle of the Mechanical Theory of Heat, Clausius writes:

> "A passage of heat from a colder to a hotter body cannot take place without compensation" (Clausius, 1879, p. 78)

This means that, while heat can spontaneously transfer from hot to cold by conduction, making the hot body cooler and the cold body hotter, heat cannot spontaneously pass from cold to hot. The phrases 'without compensation' and 'spontaneously' clearly do a lot of work here; it is possible to transfer heat from cold to hot, but work has to be done to affect this. For example, the Carnot engine run in reverse takes in work and transfers a quantity of heat from cold to hot. This work may, for example, have come from another Carnot engine which extracted work from heat moving from hot to cold. In this set-up, the transfer of heat from cold to hot was not spontaneous because work had to be done. In Clausius' wording, the transfer from cold to hot was 'compensated by' the work extract from a transfer of heat from hot to cold.

In his mathematical development of the Second Main Principle, Clausius shows[3] that it implies the existence of the state function S, which he dubs the *entropy* defined as $dS := dQ/T$. In the same way that the First Main Principle implied the

---

[3] Clausius' original argument is clear and detailed and may be found in Clausius (1879, pp. 69–107).

existence of the state function U, the Second Main Principle implies the existence of the state function S.

Myrvold (2011, p. 239) argues that this version of the second law hangs on a distinction between heat and work: it becomes false if ‘heat’ is replaced with ‘work’, since we may easily set up the situation where a piston of cold gas at high pressure does work on a piston of hot gas at low pressure. In this case, work is transferred from the hot body to the cold body ‘without compensation’. Thus the second law requires for its formulation a basic distinction between heat and work.

In the caloric theory, as already noted, this distinction is easily grounded in the fact that work is a form of energy and heat is a material substance. But in the kinetic theory, both are now forms of energy, and it is the distinction between them on which thermodynamics rests. For example, Myrvold writes:[4]

> “[T]he science of thermo-dynamics has at its core a distinction between two modes of energy transfer between physical systems: as heat, and as work.” (Myrvold, 2020)

But what grounds this distinction? Work and heat, now interpreted under the kinetic theory as forms of energy, have been brought conceptually closer together. This makes it much harder to say precisely the way in which they differ. Under the kinetic theory, as Myrvold notes, there appear to be more similarities between heat and work than there are differences:

> “On the kinetic theory of heat, when a body is heated, the total kinetic energy of its molecules is increased, so, for body A to heat body B, parts of A must interact with parts of B to change their state of motion.

---

[4] A similar view is expressed by Tribus (1961, p. xiii).

> When A does work on B, it is again the case that parts of A act on parts of B to change their state of motion." (Myrvold, 2011, p. 239)

In this way, the result of heating a body and the result of doing work on a body is in both cases a change in their state of motion. But despite this similarity, there is some difference between heat and work which we can discern. Myrvold isolates this difference as follows:

> "The difference is in heat transfer, energy is transferred to the parts of the body in a haphazard way; the resulting motions cannot be tracked. This limits our ability to recover the energy as work." (Myrvold, 2011, p. 239)

This does seem to put a finger on something about the distinction between heat and work. While work done on a system manifests itself in motion of parts of the system we 'keep track of', like the work done by pressure on a piston to increase the volume. The energy is 'useful'. By contrast the heating of a system is a transfer of motion to those parts which we cannot keep track of and which is somehow not useful. A similar view on the distinction between heat and work is given by Maxwell:

> "Available energy is energy which we can direct into any desired channel. Dissipated energy is energy we cannot lay hold of and direct at pleasure, such as the energy of the confused agitation of molecules which we call heat."

The distinction between work and heat corresponds to the distinction between available energy and dissipated energy. Maxwell clearly has anthropocentric considerations in mind as he discusses the distinction between heat and work, calling the energy of heat 'confused' and work the energy we can 'direct at pleasure'. Myrvold sums up this view on the distinction between heat and work as *means-relativity*: whether energy counts as heat or as work depends on the means we have available to us to manipulate the system and extract useful work. One way to view Maxwell's Demon is as a way of making plausible the idea that the distinction is means-relative. While we humans cannot exploit the 'confused agitation' of the molecules of the gas to extract work, the demon can and so to him, given his means, the motion of the molecules is not confused or haphazard, but systematic motion it can exploit to extract useful work. It is a small step from this view to introduce the concept of 'information' and so we step into a vast and fascinating literature on the status of information in physics.

We shall avoid that discussion, contenting ourselves with our conclusion that the distinction between heat and work on the kinetic theory of heat is seen by some as putting a kind of 'human smell' at the very foundations of the science thermodynamics.[5] This view should be taken seriously as a way of understanding what thermodynamics is about. But it is not the only view of thermodynamics' foundations which exist in the literature: we move on to consider the next characterisation.

2.1.2 The science of unobserved and observed degrees of freedom

In this section, I examine the two other characterisations of thermodynamics that exist in the literature. These characterisations stand in contrast to the first, in

---

[5] Bridgman (1961) wrote that thermodynamics "smells more of its human origin than other branches of physics".

that they make no mention of heat or work as fundamental concepts.

The second view listed in the introduction characterises thermodynamics as the study of physical systems with a very large number of degrees of freedom.[6] For example, Kubo writes:

> "Systems which are the objects of thermodynamics are macroscopic entities called systems with extensions in space and time which are accessible to normal processes of measurements. Such a system may consist of a great number of material particles (e.g. molecules, atoms, electrons, etc.) or of field quantities such as electromagnetic fields. In either case, they are dynamical systems containing an extremely large number of degrees of freedom. Systems composed of only a small number of degrees of freedom are not the object of thermodynamics." (Kubo, 1968, p. 1)

The significance of having a large number of degrees of freedom is emphasised by Callen:

> "Thermodynamics is the study of the macroscopic consequences of myriads of atomic coordinates, which, by virtue of the statistical averaging, do not appear explicitly in a macroscopic description of a system." (Callen, 1960)

This large number of degrees of freedom allows us to consider the statistical average of certain physical quantities, like the energy. To see briefly how this works,

---

[6] See, for example, Riedi (1988, p, xi), Kubo (1968, p 1) and Callen (1960). The sense in which I am using the phrase 'degrees of freedom' here is liberal; I intend it to mean some variables used to characterise the state of the system. This intended meaning distinguishes it from more narrow readings of the phrase such as a variable appearing quadratically in the Hamiltonian or as the sense used in Gibbs' Phase rule where it refers to independent intensive parameters.

take the standard textbook example of the ideal gas consisting of N particles. Let these particles be distributed over the single particle μ-space according to some distribution $f(\mathbf{x}, \mathbf{v})$ where $f(\mathbf{x}, \mathbf{v})d\mu$ is the number of particles in some infinitesimal region of μ-space. Then, the internal energy of this box of particles is the average kinetic energy of the particles: $U := \int_{-\infty}^{\infty}(1/2)m\mathbf{v}^2 f(\mathbf{x}, \mathbf{v})d\mu$.

But it is important to recognise that this view is based on the assumption that statistical mechanics is the reductive base for thermodynamics; in our example, we are explicitly invoking the microphysics of the ideal gas and the distribution of these microscopic particles in order to form the statistical average. On this view, microphysics is taken to be fundamental, with the macroscopic properties derived from statistical averaging. While the jury is still out on exactly how thermodynamics reduces to statistical mechanics, it is a very plausible assumption. But, as reasonable as it sounds, it should not be necessary to invoke statistical mechanics at the point of asking what thermodynamics is. We should be able to describe thermodynamics on its own terms, without viewing it as 'the subject which is reduced by statistical mechanics' in much the same way as we are able to describe classical mechanics on its own terms, without having to describe it as 'the subject which results from taking the classical limit of quantum theory'. If we are after a characterisation of thermodynamics which does not appeal to other theories, then this second characterisation will not do.

The third view characterises thermodynamics as the science founded on the distinction between the observed and unobserved degrees of freedom.[7] For example, Adkins writes:

---

[7] Other examples of commentators who hold this view are Dugdale (1966), Sears and Salinger (1975), Landsberg (1978, p. 1) and Uffink (2007, p. 925).

> “[A]lthough its application had at first been restricted to thermal engineering, its laws were soon recognized to be of such great generality as to be useful and important in many other branches of science also. Thermodynamics sets out to describe and correlate the directly observable properties of substances: the volume of a gas, the expansion of a wire, the polarization of a dielectric. The laws of thermodynamics enable us to interrelate the macroscopic quantities *without making any microscopic assumptions* at all. The great generality of thermodynamics is a direct consequence of this. It is precisely because it avoids microscopic theories that it is so valuable.” (Adkins, 1983, emphasis added)

This characterisation stands opposed to the second one we considered: while the second characterisation explicitly appeals to the microscopic components of the system as the basis for statistical averaging, this third characterisation explicitly ignores them, claiming that the generality and value of thermodynamics stems from its consideration of only macroscopic quantities. This approach may be seen as an effort to say what thermodynamics is on its own terms rather than appealing to statistical mechanics as the reductive base.

We must take more care, however, to state this third view as precisely as we can. To start, Adkins seems to be conflating the microscopic/macroscopic distinction between the unobserved/observed distinction. Initially, this does seem plausible, but on closer inspection is not quite right. Firstly, it depends heavily on what we mean by ‘observed’: can degrees of freedom count as observed through an electron microscope? Is there some arbitrary scale above which degrees of freedom count as macroscopic? This is the subject of a large philosophical literature and we will not enter into it here. But it is clear at least that microscopic does not necessarily imply unobserved and so these distinctions do not always match up. In the following, I will take this third characterisation of thermodynamics to be expressed in terms of the observed/unobserved distinction.

As it stands, it is rather unclear what the precise nature of the relationship between the observed and unobserved degrees of freedom is. Adkins gives the impression that the unobserved degrees of freedom are completely ignored or are irrelevant to the study of thermodynamics. In a sense this is true, but I think it would be wrong to infer from this that the unobserved plays no role whatever in thermodynamics.

As when doing any philosophy, we must take care to say what we mean by the words that we use. Therefore, to make this third characterisation clearer, and to see more precisely what it implies about the relationship between unobserved and observed physics, I will enlist the help of Burke (1985, p. 108), by paraphrasing his presentation of this characterisation of thermodynamics:

Thermodynamics is the study of thermodynamic systems. These systems are defined to be physical systems with both 'unexamined' or 'unmeasured' degrees of freedom and 'measured' degrees of freedom.[8] The unmeasured degrees of freedom might, for example, be the translational degrees of freedom of the individual atoms in an ideal gas. The measured degrees of freedom might be, for example, the mole number or volume of the ideal gas. The configuration space of the ther-

---

[8] The terms 'unexamined' and 'unmeasured' are those used by Burke and are taken by him to be synonymous. My presentation will also take these to be synonymous with each other and with 'unobserved'. There is much to be said about the relationship between measurement and observation, but I will avoid discussion of the philosophical analyses of these concepts, and assume that the measured/unmeasured and observed/unobserved distinction are the same.

modynamic system is a manifold whose points represent the thermodynamic state of the system. The manifold is given coordinates representing the measured degrees of freedom plus one representing the unmeasured degrees of freedom. This coordinate is called entropy.

Burke's exposition of thermodynamics gives us a much better idea of the different roles observed and unobserved degrees of freedom play in the foundations of thermodynamics. We begin by acknowledging that the physical system under study has some degrees of freedom. Some we can measure and observe, like volume and mole number. Others, we cannot measure and observe: they are 'hidden'. But this is not to say that we ignore them completely; we cannot, since these degrees of freedom, taken together, can still store energy which can be measured. What we can ignore, however, is the exact nature of the individual degrees of freedom. For example, the ideal gas has a very large number of degrees of freedom from the translational degrees of freedom of the individual atoms. But for the purposes of doing the thermodynamics of the ideal gas, we do not even need to know the true unobserved nature of the gas (that it is made up of atoms with translational degrees of freedom). We merely need to acknowledge that there are some unobserved degrees of freedom which can store energy, whatever they may be, and lump them together in a single degree of freedom called entropy.

We now have two main contenders for answers to our question: the first, relying on the distinction between heat and work and the third, relying on the distinction between unobserved and observed degrees of freedom. Superficially, they appear to be very different. Can we say anything more about the link between these two? There is a way to bridge this gap and it is stated[9] most concisely by Tschoegl

---

[9] The same idea is also stated by Callen (1960, p. 7)

(2000):

> “Work is transfer of energy to the macroscopically observable coordinates of motion of the constituents of matter.
>
> [...]
>
> Heat is transfer of energy to the macroscopically unobservable (or hidden) coordinates of motion.” (Tschoegl, 2000, p. 4)

We thus have a clear link between the concepts of heat/work and observed/unobserved degrees of freedom which allows us to unify these two characterisations. In this third characterisation, however, heat and work no longer appear as the primitives that they were in the first. They each receive an analysis in terms of energy transfer to the degrees of freedom of the system. We are now in a position to compare the first and third characterisations of thermodynamics.

This third characterisation brings a new perspective to the heat/work distinction: since heat is now viewed more generally as energy transfer to the unobserved degrees of freedom (the entropy), it can now be interpreted as a work term, meaning that the heat/work distinction (which views heat as an altogether different transfer of energy to work) collapses. To illustrate: suppose we have two volumes of gas separated by a piston with each volume at different pressures and volumes p_1, V_1 and p_2, V_2 where $p_1 > p_2$. If the piston is allowed to move, the pressures will equalise. More precisely, the gas with the higher pressure does work on the gas with the lower pressure and energy is transferred from the degree of freedom V_1 to the degree of freedom V_2. In an analogous way, suppose we have two volumes of gas at different temperatures T_1 and T_2 where $T_1 > T_2$ which are brought into thermal contact. Roughly speaking, the temperatures will equalise and heat will flow from gas 1 to gas 2. More precisely, the gas at T_1 does work on the gas at T_2 and energy is transferred from the entropy degree of freedom S_1 to the entropy degree of freedom S_2. Under the first characterisation of thermodynamics, these processes would have been classified as distinct; in the first case, one system does work on another; in the second case, heat flows from one system to another. Under this third characterisation, these processes are both cases of one system doing work on another. The only difference is the 'type' of degree of freedom the energy is transferred between: in the first case, energy is transferred between an 'observed' degree of freedom, the volume, while in the second case energy is transferred between the aggregate of the 'unobserved' degrees of freedom, the entropy. This perspective on things has the merit of explaining Myrvold's observation, quoted above, why heat transfer and work transfer share the feature that they both impart motion to parts of the system.

In light of this, the third characterisation allows us to reassess the role that anthropocentric considerations such as means-relativity play in the foundations of thermodynamics. In the first characterisation, they played a key role in defining the concepts of heat and work on which the formulation of the laws rests. And now, just as before, where the challenge arises to say what grounds the distinction between heat and work, so the challenge arises now to say what is the distinction between the degrees of freedom which get aggregated into the entropy variable and those which do not. In this section, we have referred to this distinction, in line with other authors, as the 'unobserved/observed' distinction and readers will be forgiven for thinking that this implies that observation, measurement or some other essentially anthropocentric considerations lie at the foundations of thermodynamics. However, this reference to observation is a little unfortunate, in that the philosophical study of observation has shown us that there is no easy way to characterise this distinction in a non-vague or non-arbitrary way. In fact, appealing to anthropocentric considerations such as observation, measurement and means is not necessary to say what the science of thermodynamics is. To do this, it will be helpful to form a direct comparison of how the first and third characterisations define thermodynamics.

The first characterisation states that thermodynamics is based on a distinction between two types of energy transfer between systems: heat and work. Adopting the anthropocentric, means-relative way of grounding this distinction, work is defined to be 'useful energy, relative to our means' and heat is defined to be 'not useful energy, relative to our means'. Heat and work, the basic foundational concepts of this characterisation, are therefore defined in anthropocentric terms. Changing the available means changes the relative proportion of energy which counts as heat or work. We therefore need to specify our available means in order to model a system in thermodynamics. But we don't need to specify the means in order to say what thermodynamics is in the first place; what defines the science thermodynamics is the fact that there is some distinction between heat and work relative to our means, whatever those means may be.

The third characterisation says that thermodynamics is based on a distinction between two types of degrees of freedom of the system under study: those which are treated together as the aggregate degree of freedom, called entropy, and those which are not. The heat transferred is defined as the work done on or by the entropy degree of freedom. While we introduce a distinction between two types of degrees of freedom, we assimilate heat into the work terms and collapse the heat/work distinction of the first characterisation. Now, a specification of which degrees of freedom are aggregated into the entropy variable is needed to create a specific model of a physical system in thermodynamics and such a specification may well depend on anthropocentric considerations such as means, observation or measurement. But, in just the same way as we did not need to specify the means in order to define thermodynamics under the first characterisation, we do not need to specify here which degrees of freedom to aggregate into the entropy variable in order to define thermodynamics. Under this third characterisation, what defines the science of thermodynamics in the first place is the fact that there is the entropy variable which aggregates some degrees of freedom, whatever those may be, and whatever considerations, anthropocentric or not, go into determining which degrees of freedom we aggregate into the entropy variable.

With these two characterisations side by side, we can see that the third characterisation 'demotes' anthropocentric considerations from the definitions of the central concepts of thermodynamics to the point of specifying a concrete thermodynamic model of a physical system. More precisely, such considerations may enter at the point of specifying the variables with which to define the state of the system.

In the following, I will adopt the third characterisation of thermodynamics on the basis that it captures the essential features of the first characterisation more generally and precisely. In the next Section, I describe the formalism of this formulation of thermodynamics.

2.2 Geometric Thermodynamics

The formalism of the third characterisation of thermodynamics is known as geometric thermodynamics. In this Section, I will present its formalism and concepts[10] adopting a postulate approach before commenting on its historical and philosophical significance.

Postulate 1. The states of the thermodynamic system are represented by points in the *configuration space*: a $(k+1)$-dimensional smooth real manifold M with global coordinates $(X_0, X_1, X_2..., X_k)$ known as *extensive variables*, where $U := X_0$ is called the *internal energy* and $S := X_1$ is called the *entropy.*

The configuration variables X_0 and X_1 are singled out as 'special' in this formalism. Indeed, we will see that the energy and the entropy are concepts which play a central role in interpreting thermodynamics and in expressing the first and second laws of thermodynamics in this formulation. The interpretation of the other variables $X_i, i > 1$ depends on the type of physical system under study, but for a simple gas may represent quantities like the volume V and mole number N. The reason for calling these variables extensive will be the subject of further discussion in the next Section.

Postulate 2. There exists a smooth function f between the extensive variables

$$U = f(S, X_2, ..., X_k) \tag{2.2.1}$$

defining a k-dimensional surface in M.

---

[10] The formalism I present largely follows Callen (1960) but the theory has been equivalently expressed using the mathematics of contact manifolds by Hermann (1973), Burke (1985), Arnold (1990) and Mrugała (1978) and Mrugała (2000), all of whom in fact cite Gibbs (1878), Gibbs (1873b), and Gibbs (1873a) as the original expression of this formulation of thermodynamics.

This postulate leaves the functional form of f wide open and does not constrain its mathematical properties. This is important because we do not wish to exclude by postulation certain physical systems from being modelled as thermodynamic. However, we will single out for discussion a certain, particularly simple, subset of thermodynamic models whose fundamental relation has the following properties:

(a) f is a first-order homogeneous function of $S, X_2, ..., X_k$:

$$f(\lambda S, X_2..., \lambda X_k) = \lambda f(S, X_2..., X_k) \tag{2.2.2}$$

for every positive real number λ and for all $S, X_2 \ldots, X_k$.

(b) U is a monotonically increasing, surjective function from S to the real numbers, all other variables held fixed.

(c) U is a convex function of its arguments. This convexity property will be discussed in more detail below.

The existence of Equation (2.2.1) which Gibbs referred to as the 'fundamental relation', expresses in 'integral form' what is more familiarly known as the *first law of thermodynamics.* To see this, we take derivatives of both sides and express the fundamental relation as

$$dU = TdS + \sum_{i=2}^{k} P_i dX_i \tag{2.2.3}$$

where the $P_i := \partial f/\partial X_i$ are called *intensive parameters* with $T := \partial f/\partial S$ called the temperature. Being partial derivatives of f, the P_i are functions of $X_1, ..., X_k$:

$$P_i = \mathcal{P}_i(X_1, ..., X_k) \tag{2.2.4}$$

Writing the intensive parameters as such functions, they are known as the *equations of the state.*

We would like to be able to identify a model of thermodynamics with the same precision as we do in classical and quantum mechanics, each of which is equipped with a precise definition of the mathematical structure of their models. For example, a model of Hamiltonian mechanics is a symplectic manifold, and a model of quantum mechanical system is a Hilbert space with an algebra of observables. However, standard texts in the philosophy and physics of thermodynamics yield no such precise definition. A thermodynamic system is sometimes vaguely characterised a 'portion of the universe' which we select for investigation (Adkins, 1983; Sears and Salinger, 1975). More common in treatises on orthodox thermodynamics is to emphasise the distinction between the system and the environment. For example, Kondepudi and Prigogine (1998, p. 4) write: "Thermodynamic description of natural processes usually begins by dividing the world into a 'system' and its 'exterior', which is the rest of the world." This is all very well, but it does not give us a precise characterisation of the subject's target of study.

Slightly more enlightenment is achieved by commentators on geometric thermodynamics and related formulations. Callen (1960, p. 8) defines a *simple system* as a physical system which is "macroscopically homogeneous, isotropic, uncharged, and chemically inert, that are sufficiently large that surface effects can be neglected, and that are not acted on by electric, magnetic, or gravitational fields." Callen's definition introduces the property of being macroscopically homogeneous as an essential property of thermodynamic systems. This is certainly a step towards greater conceptual precision, but is not yet adequately formalised.

Tisza (1961, p. 7) defines a *Thermodynamic Simple System* to be “a finite region in space specified by a set of [configuration] variables $X_0, X_1, ..., X_k$.” Tisza's definition introduces the idea that the state of a thermodynamic system is specified by the $k+1$ extensive configuration variables. This aspect of the definition is also a step in the right direction, but does not include homogeneity of the thermodynamic system, something Tisza himself emphasised elsewhere (Tisza, 1961, p. 23).

Perhaps the most precise definition is to be found in the formulation of geometric thermodynamics in the language of contact geometry[11]. To briefly introduce it: begin by defining the *contact manifold* as the pair (N, θ). Here, N is the *thermodynamic phase space*, a $(2k+1)$-dimensional space, the ‘contact bundle’, constructed out of the $(k+1)$-dimensional configuration space M by taking the union of the cotangent spaces at every point $p \in M$. θ is the so-called *contact form*: $\theta = dU + TdS + \sum_{i=2}^{k} P_i X_i$. Now let $\mathcal{I}$ be a k-dimensional submanifold of N and $\varphi : \mathcal{I} \to N$ the imbedding of $\mathcal{I}$ into N. We say that $\mathcal{I}$ is an *integral submanifold* if $\varphi^* \theta = 0$. Finally, we say that a thermodynamic system is represented in the contact manifold by the k-dimensional integral submanifold of $\theta = 0$.

This definition is extremely rich and clearly requires a lot of mathematical structure. We shall not delve into the intricate details of contact geometry, since the full-blown machinery is not necessary for our purposes. The point of briefly introducing it is to demonstrate that it is possible to achieve a characterisation of thermodynamic systems to the same degree of precision as our characterisation of classical and quantum systems. In the following, we shall take this definition of

---

[11] See Arnold (1990), Mrugała (2000), Mrugała (1978), Burke (1985), and Hermann (1973).

thermodynamic system in terms of contact geometry as inspiration, simplifying it greatly while retaining its essential content. Based on Posulates 1 and 2, we define a thermodynamic system as follows.

Thermodynamic System. A *thermodynamic system* is the pair (M, f), where M is a $(k+1)$-dimensional manifold with global coordinates $U, S, X_2, ..., X_k$ and $U = f(S, X_2, ..., X_k)$ where f is the fundamental relation.

It shall also be necessary to give a precise definition of *composite system* and *subsystem.* While many commentators on thermal physics seem to deem it necessary to draw attention to the distinction between composite system and subsystem, this relationship is rarely made precise. For example, Callen (1960, p. 23) writes: "Given two or more simple systems, they may be considered as constituting a single *composite system.*" He does not spell out exactly how they may be considered as such, and does not say precisely what the mathematical relationship is between the composite system and the subsystem.

Tisza (1961) recognises the problem of stating this relationship precisely:

> "We have mentioned that the key to the continued development of thermodynamics lies in those assumptions that are usually not explicitly mentioned. Among the most significant of these is the assumption of a *constructive principle.*" (Tisza, 1961, p. 6)

By 'constructive principle', Tisza means some way of building a thermodynamic system (the composite system) out of other thermodynamic systems (the subsystems). Thus, in addition to his definition of a thermodynamic simple system, given above, he also introduces the following definition of composite system: "A conjunction of spatially disjoint simple systems. A composite system is obtained by uniting separate systems or by partitioning a single system. The simple systems that form the composite systems are called its subsystems" (Tisza, 1961, p 7). This is still lacking in the mathematical and formal precision required for this concept to be useful; what does it mean to 'unite' systems into a composite? And what does it mean to 'partition' a simple system into subsystems? Tisza provides the answer in the form of the *Additivity postulate*: "In a composite system $X_i = \sum_a X_i^{(a)}$ where the summation is over the subsystems. The quantity on the left belongs to the composite system." (Tisza, 1961, p. 8). With this postulate, Tisza has provided a precise way in which we may understand the relationship between the subsystems and their composite.

Tisza's definition works in tandem with his definition of thermodynamic system. Since my thermodynamic system is different from Tisza's, I will need to propose a definition of composite system which works in tandem with my thermodynamic system definition. Tisza's definition essentially defines the variables of the composite system as a specific function of the variables of the subsystems (add them). Mine will also do this, but departs from Tisza's in two significant ways. First, I generalise Tisza's idea by allowing the variables of the composite to be any function of, not just the sum of, the variables of the subsystems[12]. Second, I specify explicitly the need for the fundamental relation of the composite system to satisfy Postulate 2.

Suppose we have k thermodynamic models $\mathcal{M}^{(1)}, \ldots, \mathcal{M}^{(k)}$, whose manifolds

---

[12] Looking forward, my reason for generalising Tisza's definition in this way is to help us understand thermodynamic mixtures. I will propose that we should endeavour to understand mixtures as composite systems where the subsystems are the individual components of the mixture. The mixture's configuration variables, I will argue, are not simply sums of variables of the individual subsystems.

are each given coordinates $(U^{(i)}, V^{(i)}, N^{(i)})$ for each $i = 1, ..., k$. I propose to define the composite of these individual thermodynamic models as follows.

Composite System. A *composite system* is the $(n+1)$-tuple $(\mathcal{M}^{(1)}, \ldots, \mathcal{M}^{(n)}, \mathcal{M})$ where $\mathcal{M}^{(1)}, \ldots, \mathcal{M}^{(n)}$ are n thermodynamic models whose manifolds $M^{(i)}$ may be given global coordinates $X_1^{(i)}, \ldots, X_{\dim(M^{(i)})}^{(i)}$ for $i = 1, \ldots, n$ and $\mathcal{M}$ is a thermodynamic model whose manifold M can be given global coordinates $X_0, X_1, \ldots, X_m$ each of which is a function of the coordinates $X_1^{(i)}, \ldots, X_{\dim(M^{(i)})}^{(i)}$ for all i. For each $i = 1, \ldots, n$, we say that $\mathcal{M}^{(i)}$ is a *subsystem* of the composite system.

Specifying that the coordinates of the composite system must be functions of the coordinates of the subsystems incorporates an important assumption about the relation between the composite system and its subsystems, namely that the composite system supervenes on its subsystems: there may be no change in the composite system without some change in the subsystems. I do not claim that this is the only definition of composite system possible—there may well be others—but I do claim that it relies on a plausible thesis about the relation between composite systems and their subsystems. It is also essential that the fundamental relation f of the composite system satisfy Postulate 2, for otherwise (M, f) would not be a thermodynamic system.

My justification for pursuing a precise definition of composite system is, in fact, the very same as Tisza's (Tisza, 1961, p. 12): to be able to state the second law of thermodynamics. It is not supposed to be obvious that the very statement of the second law depends on a precise relationship between a composite system and its subsystems. But in Section 2.2.4 I will argue that it is the case.

With these definitions and postulates in place, we will now proceed to discuss them and explain the physical and philosophical significance of this framework.

2.2.1 Postulate 1: Energy and Entropy

In orthodox thermodynamics, energy and entropy are given starring roles in its foundations through the first and second laws which imply the existence of the state functions energy and entropy respectively. In the geometric formulation, energy and entropy also take on foundational roles. This is clear in Postulate 1, in which we see that the interpretation of the configuration variables X_0 and X_1 is always energy and entropy, respectively. In this section, we discuss how the precise nature of the roles of energy and entropy in the geometric formulation parts in two significant ways from the nature of their roles in orthodox thermodynamics.

The first way in which the geometric and orthodox formulations part is in the fact that energy and entropy are treated as conceptually fundamental in the geometric formulation while they are derived in the orthodox formulation. To see this, recall from our discussion of Clausius that the existence of the internal energy U is a consequence of the First Main Principle: assuming heat and work as two fundamental modes of energy transfer, together with the Principle, implies that there exists this state function U. Regarding entropy, recall our discussion above that the existence of the fundamental relation which gives the energy as a function of the other configuration variables implies a version of the more familiar first law of thermodynamics; if we take the differential of the fundamental relation, we get Equation (2.2.3): $dU = TdS + \sum_i P_i dX_i$. Compare this with Equation (2.1.1): $dU = dQ + dW$. Notice that in the orthodox first law, heat is a fundamental concept while the entropy S is derived: in Clausius' mathematical development of his Second Main Principle, he shows that the state function entropy exists while taking heat and temperature[13] as given. This contrasts with Equation (2.2.3) where entropy is taken to be fundamental and temperature is defined to be the derivative of the energy with respect to the entropy. Heat has so far made no appearance in the geometric formulation of thermodynamics, and actually requires more mathematical structure than has so far been presented in order for it to be defined.

One might wonder whether these assumptions, rather than derivations, of the existence of energy and entropy functions count as a point against adopting the geometric formulation of thermodynamics; while the existence of the entropy function is simply postulated in the geometric formulation, great care is taken in the orthodox formulation to derive it. This could potentially lead to Russell-type accusations against the geometric formulation of theft over honest toil. However, this is not entirely fair. In the geometric formulation, entropy is not simply introduced without further comment: considerable effort is expended in justifying its inclusion as one of the configuration variables with a special interpretation: an aggregate variable of the unobserved degrees of freedom. Furthermore, entropy being a derived concept in orthodox thermodynamics does not necessarily make it a 'better' theory. Different formulations of the same science are different theories with the ability to ask and answer different types of questions.

This point of contrast between the orthodox and geometric formulation resurfaces as another point of difference between the theories: in the orthodox formulation, *processes* (coming in the various flavours of cyclic, reversible, irreversible and quasistatic) between states take on fundamental roles in the development in the

---

[13] Indeed, even the existence of absolute temperature depends on the Second Main Principle.

theory; the existence of the energy and entropy functions could not have been done without them. By contrast, the geometric formulation places the foundational emphasis squarely on *states*; by assuming the existence of energy and entropy from the outset, we are able to immediately characterise the state of the system; processes play no foundational role here. It is for this reason that this formulation is sometimes known as *thermostatics*.

Now, one might well wonder about the relative merits of these two formulations. If one is more operationally minded, then perhaps the orthodox version appeals, given its meticulous derivation of energy and entropy in terms of more operationally understood notions of heat and work. On the other hand, if one is perhaps more axiomatically minded, the geometric formulation might appeal. Assuming the formulations are empirically equivalent[14] one's choice between them becomes partly a matter of taste, and I choose to adopt the geometric formulation of thermodynamics in this thesis. Such matters of taste may be philosophical and concern the well-definedness of certain concepts. Norton (2016) discusses the the notion of a reversible process in thermodynamics and argues that standard expositions of the concept are internally contradictory. Norton argues that this defect can be remedied by giving a definition of the concept which does make sense and is able to ground all standard thermodynamic reasoning. If one is worried by the notion of thermodynamic reversibility, this might give one reason to adopt thermostatics. On the other hand, one might be worried by the conceptual priority given to energy and entropy in thermostatics, in which case one may choose to adopt the orthodox formulation.

---

[14] I see no reason why they should not be so. I know of no phenomenon which is captured by one theory and not by the other, nor am I aware of any discussion on this matter in the literature.

The second way in which the formulations part regards the widespread interpretation of the orthodox first law as a special case of the energy conservation principle. We must be careful here, because energy conservation is a principle from mechanics and does not obviously carry over into thermodynamics. In mechanics, energy conservation is the claim that the hamiltonian h is invariant under the phase flow map φ_t associated with the vector field H^a generated by the hamiltonian. Stating this principle in hamiltonian mechanics therefore requires the symplectic structure of the phase space and an interpretation of the phase flow as 'flow through time'. Thermodynamics does not have this structure and we have not even defined the notion of 'flow' or 'process'. It is therefore not helpful to think of the first law of thermodynamics as energy conservation in this sense. Perhaps a more helpful way of conceptualising the fundamental relation is as giving the energy of the thermodynamic system as a function of its degrees of freedom, in just the same way as the Hamiltonian gives the energy of a mechanical system as a function of its degrees of freedom. Seen this way, the existence of the fundamental relation is the claim that the energy of the thermodynamic system is determined entirely by the values of the thermodynamic system's degrees of freedom which may, at a stretch perhaps, be interpreted as some kind of energy conservation.

Adopting this interpretation of the first law, we gain an insight into the role of the configuration variables as the system's degrees of freedom and hence implies an interesting interpretation of the entropy: U is a function of S, and if U is to be interpreted as a function of the degrees of freedom, then entropy is a degree of freedom. Interpretations of entropy, both thermodynamic and statistical mechanical, abound in an extremely large literature involving entropy's connection with concepts such as time, information and disorder. Compared with these, the interpretation of entropy suggested here is very little discussed, even among adherents of the geometric formulation. The clearest statement of this interpretation is by Burke (1985) which I have already been briefly mentioned in Section 2.1.2. I shall expand on this interpretation a little more here.

Entropy interpreted as a degree of freedom of a thermodynamic system is a crucial assumption in the foundations of thermodynamics. As is clear from the postulates above, every thermodynamic system has some number of degrees of freedom (the configuration variables) the particular nature of which depend on the kind of physical system under study but one of them is always entropy: the energy of the ideal gas is a function of S, V, N where V is the volume and N is the mole number, the energy of a magnetic thermodynamic system is a function of S, B where B is the magnetic field, the energy of a liquid surface is a function of S, A where A is the surface area. We may say that what makes a model of a system 'thermodynamic' is the inclusion of the entropy as a 'special' degree of freedom distinct from volume, mole number and the like. But what exactly is this distinction? The entropy is the aggregate of all the 'hidden' degrees of freedom which are not directly measurable. The simple ideal gas is the perfect example: a gas of N monatomic particles has $3N$ translational degrees of freedom. These are not individually modelled in thermodynamics, but they still can store measurable energy, just as the volume degree of freedom can. We project all these translational degrees of freedom on to one aggregate thermodynamic degree of freedom, the entropy, acknowledging that energy may flow into and out of these degrees of freedom, but ignoring their precise nature. What makes thermodynamics so powerful and general is its ability to acknowledge that physical systems have certain degrees of freedom which can store energy but which may not be accessible or even known. Through the entropy degree of freedom, thermodynamics has the ability to model physical systems where we do not have access to some of its physics or where we choose to deliberately ignore the nature of some of its degrees of freedom.

Postulates 1 and 2 are minimally sufficient to characterise a thermodynamic system. That is, is is sufficient for a physical system to be modelled as thermodynamic that it has an entropy degree of freedom and that there be a fundamental relation defined a k-dimensional surface in the $(k+1)$-dimensional space. However, as already mentioned, Postulate 2 leaves the mathematical properties of f completely unspecified in order that no physical system is excluded by postulation from being modelled as thermodynamic. But, for now, we may consider a particularly simple class of thermodynamic models whose fundamental relation satisfies the properties of homogeneity, monotonicity and convexity. This serves not only to highlight the physical systems under consideration in this thesis but also serves to bring our attention to thermodynamic systems where these assumptions do not apply.

2.2.2 Postulate 2 (a): Homogeneity

The homogeneity postulate may be interpreted as stating that thermodynamic systems have a scaling symmetry, capturing the idea that physical systems modelled as thermodynamic 'look the same' on all scales, that is, they are homogeneous. The postulate also supplies the reason for calling the configuration variables extensive. To see this more precisely, take the configuration variables $\{U, S, X_2 ..., X_k\}$ defining the state of the thermodynamic system and define a new set obtained by multiplying each by some positive real number λ: $\{\lambda U, \lambda S, \lambda X_2 ..., \lambda X_k\}$. If we interpret the variables $\{U, S, X_2 ..., X_k\}$ to be necessary and sufficient for characterising the 'quantity' of a substance, then there is a sense in which the state of the system defined by the variables $\{\lambda U, \lambda S, \lambda X_2 ..., \lambda X_k\}$ is 'like' the system defined by variables $\{U, S, ..., X_k\}$ and yet also 'different': the variables correspond to a different quantity of the same substance, related to the previous quantity by the scaling parameter λ. The precise sense in which this new system is 'like' the old one is if the fundamental relation is invariant under the transformation $X_i \to \lambda X_i$. Applying this transformation, Equation (2.2.1) becomes

$$\lambda U = f(\lambda S, \lambda X_2 ..., \lambda X_k). \tag{2.2.5}$$

If f is a first-order homogeneous function of the variables $S, X_2 ..., X_k$, i.e. if it satisfies the homogeneity assumption of Equation (2.2.2), then we have

$$\lambda U = f(\lambda S, \lambda X_2 ..., \lambda X_k) = \lambda f(S, X_2 ..., X_k) \implies U = f(S, ..., X_k). \tag{2.2.6}$$

That is, if f is a first-order homogeneous function of the variables $S, X_2 ..., X_k$, the scaling transformation $X_i \to \lambda X_i$ is a symmetry of the fundamental relation. It is in this mathematical sense that thermodynamic systems are homogeneous.

This interpretation of 2(a) gives us some much needed insight into the alternative nomenclature of the configuration variables as 'extensive' stated in Postulate 1. Extensive variables are typically characterised in two ways. The first way states that extensive variables are those which scale with the size of the system. For example, Kondepudi and Prigogine (1998) write that the variables "such as volume and mole number, which are proportional to the size of the system, are called extensive variables." The second way describes extensive variables as those which are proportional to the the mass. For example, Kestin (1971, p. 8) writes: "A property is called extensive if its value in such a system is proportional to mass." and Sears and Salinger (1975, p. 3) write: "Those properties of a system in a given state that are proportional to the mass of a system are called *extensive.*" However, these characterisations of extensivity are not adequate. The first way is ambiguous: what is meant by the system's size? Is it the volume? Mole number? Both? The second way is not sufficiently general: mass is not guaranteed to be a variable that characterises a thermodynamic system. Plausibly it might be one of the variables used in defining the state of a gas where we convert the volume into mass by multiplying by the density or convert the mole number into mass by multiplying by the mass per particle. But what if the thermodynamic system's state is not characterised by mole number or volume?

In the geometric formulation and on this interpretation of 2(a) we gain a precise understanding of what it means for a variable to be extensive. They are simply the configuration variables in terms of which the fundamental relation is defined and whose scaling preserves the functional form of the fundamental relation. This way, we avoid ambiguity in the word 'size' and we also avoid 'brute fact' claims often seen in the discussion of extensivity that volume and mole number are 'obviously' extensive. Extensivity, as characterised here, is a formal notion, independent of what physical interpretation the configuration variables have.

Not all physical systems which we may wish to represent as thermodynamic models 'look the same on all scales', however. One salient example is systems in which more than one phase of the substance exists. Such systems and phase transitions themselves are modelled in thermodynamics and are precisely the cases in which the homogeneity assumption breaks down. Although such systems are not explicitly my concern in this thesis, it is worth noting briefly how the conceptual connection between the assumption of homogeneity and phase transitions comes about. Callen (1960, p. 136) shows that any homogeneous fundamental equation is consistent with certain stability criteria, and that when external constraints on certain variables change their values such that they begin to conflict with the stability criteria, phase transitions occur.

2.2.3 Postulate 2 (b): Monotonicity

The monotonicity expressed in 2(b) implies that the fundamental relation is invertible: Equation (2.2.1) may be written as $S = g(U, X_2, ..., X_k)$ where the function g results from rearranging Equation (2.2.1) to make S the subject. To prove this, it suffices to show that f is injective. For then $f : S \to \mathbb{R}$ is a bijection and thus has an inverse. The injective property of f follows immediately from the assumption that U is a strictly monotonically increasing function of S. This property is expressed as follows. $f : X \to Y$ is a strictly monotonically increasing function iff: if $x > y$ then $f(x) > f(y)$. It follows that if $x \neq y$ then $f(x) \neq f(y)$ which is exactly the injective property.

If the fundamental equation is written with U the subject then we say that it is expressed in the *energy representation*; if it is written with S the subject then it is the *entropy representation*. 2(b) also implies that $\partial U/\partial S > 0$ which, together with the definition $T := \partial U/\partial S$ is the claim that temperature is non-negative.

Some fundamental relations may not be monotonic. These, then, admit the possibility of negative absolute temperatures, in the sense that adding energy to the system will make the entropy decrease. A standard example of this phenomenon is population inversion in laser physics. Nuclear spin systems cited by Ramsey (1956) are early examples of physical systems exhibiting negative temperature behaviour. Whilst the vast majority of physical systems will have monotonic fundamental relations when modelled in thermodynamics, we would not want to exclude those systems which exhibit negative temperature behaviour.

2.2.4 Postulate 2 (c): Convexity

The convexity property of the fundamental relation in the energy representation[15] is the home of the second law of thermodynamics in the geometric formulation. In detail, the convexity property means the following:

Convexity. The function $f(X_1, ..., X_k)$ is convex iff, for all i and for any values $X_i', X_i'' \geqslant 0$ for all i and for any $\alpha \in (0, 1)$:

$$f(\alpha X_1' + (1-\alpha)X_1'', ..., \alpha X_k' + (1-\alpha)X_k'') \leqslant \alpha f(X_1', ..., X_k') + (1-\alpha) f(X_1'', ..., X_k'') \tag{2.2.7}$$

To understand the sense in which this postulate encodes the second law, I will employ the help of Wightman (1979) who attempts to demonstrate just such a claim. But before we even delve into Wightman's argument, we must draw attention to the inevitable question: which second law? A cursory survey of the literature in the foundations of thermal physics reveals a great number of versions

---

[15] If the fundamental relation is inverted to express the postulates in the entropy representation then 2(c) would be expressed as the assumption that S is a concave function of its arguments.

of the law[16]. Here, and without further discussion, I am going to adopt the version of the law adopted by Gibbs and subsequent authors adopting his approach. Roughly speaking, this 'Gibbsian' version of the law states that the values of the extensive variables of subsystems of a composite system, in the absence of any internal constraints, are those which minimise the energy[17] of the composite system. Before establishing the link between convexity and the second law, I will spell out the Gibbsian version of the law in more detail in order to understand fully what it says and how it works.

For example, suppose we have a box of gas (the composite system) divided up into two halves (the subsystems) of different pressures, separated by an immovable piston (the internal constraint). Question: what happens to the volumes when the piston is allowed to move? The intuitive picture we have in our minds is of the piston moving until the two halves of the box have reached equilibrium and the pressures have equalised. This intuition can be formulated as a precise consequence of the second law; the volumes of the two halves of the box which are attained after the removal of the constraint on the volume may be calculated by minimising the energy with respect to the constraint that the two volumes sum to the volume of the entire container. Because of this, the second law may be seen as a definition of equilibrium: we say that the system is in equilibrium (with respect to volume) when the the volumes of the two halves take on values which minimise the energy. Let us now state the second law with full precision.

Consider a thermodynamic system with extensive variables $X_0, ..., X_k$. The

---

[16] The locus classicus for this subject is Uffink (2001) in which every conceivable version of the second law is discussed.

[17] Or, if we are in the entropy representation, maximise the entropy. It is interesting to note that the second law of thermodynamics in this formulation is just as much to do with energy as it is to do with entropy, if we assume the monotonicity of the fundamental relation.

first key step is to imagine this system to be composed of two subsystems such that each subsystem (i) has variables $X_K^{(i)}$ such that $X_K^{(i)} = \lambda^{(i)} X_K$ for $i = 1, 2$ with $\lambda^{(1)} + \lambda^{(2)} = 1$. This implies that for some K, the $X_K^{(i)}$ for $i = 1, 2$ may take on any values consistent with $X_K^{(1)} + X_K^{(2)} = X_K$ (Note that this is exactly Tisza's definition of composite system, but only a special case of the definition I gave above). This particular decomposition of the system into two subsystems is sufficient for us to say what it means for the system to be in equilibrium[18].

K-**Equilibrium.** Let (M, f) be a thermodynamic system with variables $X_0, ..., X_k$ which is composed of n subsystems $(M^{(i)}, f^{(i)})$ for $i = 1, 2$ such that $X_K^{(1)} + X_K^{(2)} = X_K$ for all $K = 0, ..., k$. Then the values of the variables $X_K^{(i)}$ which are attained on the removal of the internal constraint are those such that the energy of the composite system satisfies $U(X_0, ..., X_k) = \inf_{X_K^{(1)} + X_K^{(2)} = X_K} \{U^{(1)}(X_0^{(1)}, ..., X_k^{(1)}) + U^{(2)}(X_0^{(2)}, ..., X_k^{(2)})\}$. If this condition on the subsystems' extensive quantities is satisfied we say that the composite system is in equilibrium with respect to the variables X_K, or more briefly, in K-equilibrium.

Note that the definition is indexed to an extensive variable X_K; we may choose to minimise the energy with respect to any number of the extensive variables,

---

[18] Uffink (2001, p. 360) notes that, strictly speaking, the following criterion leaves the precise nature of the equilibrium undetermined. What this condition actually gives us is the conditions under which the entropy is at an *extremum*, i.e. $dS = 0$. This leaves open the possibilities that $d^2S < 0$ and $d^2S > 0$, in which case we have entropy at a *maximum* (stable equilibrium) or *minimum* (unstable equilibrium) respectively. The system may also exist at 'local' maxima which are not the global maximum. Such states are known as *metastable*; supercooled water, which remains liquid below its freezing point, is an example of a system in a metastable state. These considerations indicate that this criterion is not one of equilibrium *tout court*. This obviously limits the scope of the convexity assumption vis-à-vis equilibrium. But this doesn't necessarily count against this formulation of thermodynamics and its associated representation of equilibrium and the second law. If anything, it is a virtue of this formulation that such distinctions and categorisations are able to be made precisely and that it is able to point clearly at further avenues at conceptual-philosophical research.

each of which yields a particular kind of equilibrium. In our example above, we removed a constraint on the volume, allowing the volume to 'flow' between the subsystems resulting in mechanical equilibrium. Again, if we choose to minimise the energy with respect to S, we get a system in thermal equilibrium. Minimising U with respect to V yields mechanical equilibrium and with respect to N, chemical equilibrium. In the special case where the system is in equilibrium with respect to all extensive variables then we make the following definition:

Thermodynamic Equilibrium. A composite system is in *thermodynamic equilibrium* if and only if the system is in K-equilibrium for all K.

With this development of the second law, we can clearly see the important role played by a precise notion of composite system and subsystem. The second law requires such a notion in order to express it; the law is about the flow of extensive quantities between subsystems of a composite. To my knowledge, the first person to be explicit about the importance of a composite/subsystem definition for expressing the second law was Tisza (1961):

> "We turn now to a methodological problem connected with the fact that the [geometric] formalism deals only with states of equilibrium while nontrivial results can be obtained only if processes are considered at least to a limited extent. The nature of this problem is illustrated by the following paradox: How are we to give a precise meaning to the statement that entropy tends toward a maximum, whereas entropy is defined only for systems in equilibrium? Thus in an isolated simple system the entropy is constant, if it is defined at all.

> This difficulty is resolved in a natural way by the artifice of composite systems that enables us to deal with more or less constrained equilibria." (Tisza, 1961, pp. 19–20)

The way I am reading this 'methodological problem' is as follows: since an entropy can only be assigned to equilibrium states, and entropy is already at an extremum in equilibrium states (per the definition of K-Equilibrium), how can we make sense of talk of entropy changes as a result of the removal of constraints?

The resolution of the apparent paradox is to argue that, while the composite system is not in equilibrium before the constraints are removed, each of the subsystems is in equilibrium. Thus each has a well defined entropy and thus, by additivity, so does the composite. We therefore avoid the paradoxical conclusion that when the entropy is defined, it must already be at an extremum; with the artifice of composite systems, we can have a well-defined entropy which is not at an extremum. We can therefore determine the final equilibrium state by finding the values of the subsystems' extensive variables which extremise the composite system's entropy. We may imagine the extensive quantities "flowing" between the subsystems until the entropy is at an extremum. The beauty of this formulation of equilibrium is that it uses the notion of composite and subsystems to analyse a non-equilibrium concept in equilibrium terms.

However, it is also important to note what the puzzle is not, as I see it. Tisza uses the phrase "tends toward a maximum" which is heavily process-laden language, invoking the notion of changes of physical quantities with respect to time. Thus, it might appear that the puzzle is to give a precise meaning to this notion of the change of entropy in processes where no state in that process is an equilibrium state, and thus, strictly speaking, no such state has a well-defined entropy. I do not see that as the puzzle here, as I hope to have already indicated in my response to Tisza in the previous paragraph. But if one does see this as the puzzle, I acknowledge that it has no solution in this ‘state-fundamental’ formulation of thermodynamics. Presumably, if one does want to make sense of such processes, some kind of non-equilibrium thermodynamic theory would be necessary.

We are now in a position to understand the second law’s connection with the convexity of the energy function. I will show, following Wightman (1979, pp. xiv–xvi), that convexity implies and is implied by the second law. Let us begin with the convexity of the energy function, stated again here for convenience:

Convexity. The function $f(X_1, ..., X_k)$ is convex iff, for all i and for any values $X_i', X_i'' \geqslant 0$ and for any $\alpha \in (0, 1)$:

$$f(\alpha X_1' + (1-\alpha)X_1'', ..., \alpha X_k' + (1-\alpha)X_k'') \leqslant \alpha f(X_1', ..., X_k') + (1-\alpha) f(X_1'', ..., X_k'') \tag{2.2.8}$$

Wightman states that the function f is convex if and only if it satisfies

$$f(X_1, ..., X_k) = \inf \{ f(X_1', ..., X_k') + f(X_1'', ..., X_k'') \} \tag{2.2.9}$$

such that $X_i = X_i' + X_i''$. This last condition on f is the second law in the energy representation if we interpret $f(X_1, ..., X_k)$ as the energy U of the composite system and $f(X_1', ..., X_k')$ and $f(X_1'', ..., X_k'')$ are the energies of the subsystems. In words, it states that if we form a set of values of the sum of the subsystems’ energies constructed by running through all possible values of the subsystems’ extensive quantities X_i', X_i'' such that they sum to some constant $X_i = X_i' + X_i''$, then the energy of the composite system (with variables $X_i, \ldots, X_k$) at equilibrium (with all internal constraints removed) is given by the infimum of that set.

Wightman leaves the details of proving this proposition to the reader. In the following, we will fill in the details. We begin by outlining some definitions and and lemmas before proving Wightman's proposition.

Definition 1. A function $f : \mathbb{R}^n \to \mathbb{R}$ is *homogeneous* if for all $x \in \mathbb{R}^n$, $f(\lambda x) = \lambda f(x)$ for any $\lambda > 0$.

Definition 2. A function $f : \mathbb{R}^n \to \mathbb{R}$ is *convex* if $f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha) f(x_2)$ for all $x_1, x_2 \in \mathbb{R}^n$ and $\alpha \in (0, 1)$.

Definition 3. A function $f : \mathbb{R}^n \to \mathbb{R}$ is *subadditive* if $f(x_1 + x_2) \leq f(x_1) + f(x_2)$ for all $x_1, x_2 \in \mathbb{R}^n$.

Definition 4. $C \subseteq \mathbb{R}^n$ is a *cone* if it is closed under positive scalar multiplication, i.e. for all $x \in \mathbb{R}^n$, $\lambda x \in C$ for $\lambda > 0$.

Definition 5. $C \subseteq \mathbb{R}^n$ is *convex* if $(1-\lambda)x_1 + \lambda x_2 \in C$ for all $x_1, x_2 \in C$ and $\lambda \in (0, 1)$.

Definition 6. The *epigraph* of a function $f : \mathbb{R}^n \to \mathbb{R}$ is the set $\text{epi}(f) := \{(x, \mu) \in \mathbb{R}^n \times \mathbb{R} : f(x) \leq \mu\}$.

We may imagine the epigraph of a function to be the set of points lying above the line of the graph.

In Lemma 1, we establish a link between the convexity of a function f and the convexity of the set which is the epigraph of f.

Lemma 1. epi(f) is convex iff $f : \mathbb{R}^n \to \mathbb{R}$ is convex.

Proof. $\to$. Let $(x_1, f(x_1)), (x_2, f(x_2)) \in$ epi(f). Assume epi(f) is convex. Then $(\lambda x_1 + (1-\lambda)x_2, \lambda f(x_1) + (1-\lambda)f(x_2)) \in$ epi(f). By definition of epi(f), $f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2)$, so f is convex.

$\leftarrow$. Assume f is convex and let $(x_1, \mu_1), (x_2, \mu_2) \in$ epi(f). Then we have $f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2)$ and $f(x_1) \leq \mu_1$ and $f(x_2) \leq \mu_2$. Therefore, $f(x_1)+f(x_2) \leq \mu_1+\mu_2$. Therefore, $f(\lambda x_1+(1-\lambda)x_2) \leq \lambda\mu_1+(1-\lambda)\mu_2$. Therefore, $(\lambda x_1 + (1-\lambda)x_2, \lambda\mu_1 + (1-\lambda)\mu_2) \in$ epi(f). So epi(f) is convex. □

In Lemma 2, we establish a link between the homogeneity of f and the conic property of the epigraph of f

Lemma 2. epi(f) is a cone iff f is homogeneous.

Proof. $\to$. Assume epi(f) is a cone. Therefore, since $(x, f(x)) \in$ epi(f), $(\lambda x, \lambda f(x)) \in$ epi(f). This means $f(\lambda x) \leq \lambda f(x)$. Also, since $\mathbb{R}^n$ is a cone, $\lambda x \in \mathbb{R}^n$ so that $(\lambda x, f(\lambda x)) \in$ epi(f). Now, since epi(f) is a cone by assumption, $(x, f(\lambda x)/\lambda) \in$ epi(f), so $f(x) \leq f(\lambda x)/\lambda$. So $\lambda f(x) \leq f(\lambda x)$. Therefore, $f(\lambda x) = \lambda f(x)$ so f is homogeneous.

$\leftarrow$. Assume f is homogeneous: $\lambda f(x) = f(\lambda x)$. Let $(x, \mu) \in$ epi(f). This means that $f(x) \leq \mu$. Therefore $\lambda f(x) \leq \lambda\mu$. By homogeneity, $f(\lambda x) \leq \lambda\mu$. Therefore $(\lambda x, \lambda\mu) \in$ epi(f). So epi(f) is a cone. □

Together, Lemmas 1 and 2 establish a link between ‘function talk’ and ‘set talk’. Thus the convexity and homogeneity of the function f have been translated into certain properties of a set related to f (its epigraph). This is important, since the proposition we want to prove is most easily done using the properties of sets, rather than functions. In particular, Lemmas 3 and 4 are easily proved using the relevant properties of sets.

Lemma 3. Let $C \subseteq \mathbb{R}^n$ be a cone. Then C is convex iff it is closed under addition.

Proof. $\rightarrow$. Let $x_1, x_2 \in C$ and assume C is convex. Then $\frac{1}{2}x_1 + \frac{1}{2}x_2 \in C$ and, since C is a cone, $x_1 + x_2 \in C$. So C is closed under addition.

$\leftarrow$. Let $x_1, x_2 \in C$ and assume C is close under addition. Since C is a cone, $(1-\lambda)x_1, \lambda x_2 \in C$, and by closure, $(1-\lambda)x_1 + \lambda x_2 \in C$, so C is convex. □

Lemma 4. $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is subadditive iff $\text{epi}(f)$ is closed under addition.

Proof. $\rightarrow$. Assume f is subadditive. Then $f(x_1 + x_2) \leq f(x_1) + f(x_2)$ for all $x_1, x_2 \in \mathbb{R}^n$. Let $(x_1, \mu_1), (x_2, \mu_2) \in \text{epi}(f)$. Then, $f(x_1) \leq \mu_1$ and $f(x_2) \leq \mu_2$. It follows that $f(x_1) + f(x_2) \leq \mu_1 + \mu_2$. Since f is subadditive, it follows that $f(x_1 + x_2) \leq \mu_1 + \mu_2$. This means $(x_1 + x_2, \mu_1 + \mu_2) \in \text{epi}(f)$. Therefore, $\text{epi}(f)$ is closed under addition.

$\leftarrow$. Assume $\text{epi}(f)$ is closed under addition and note that $(x_1, f(x_1)), (x_2, f(x_2)) \in \text{epi}(f)$. By closure, we have $(x_1 + x_2, f(x_1) + f(x_2)) \in \text{epi}(f)$. This just means that $f(x_1 + x_2) \leq f(x_1) + f(x_2)$. So f is subadditive. □

Lemmas 3 and 4 establish a link between convexity and subaddtivity. All the lemmas together combine to prove the following theorem.

Theorem 1. Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be homogeneous. Then f is convex iff f is subadditive.

Proof. $\rightarrow$. f is assumed to be homogeneous. So by Lemma 2, $\text{epi}(f)$ is a cone. Assume f is convex. By Lemma 1 $\text{epi}(f)$ is convex. Then, by Lemma 3, $\text{epi}(f)$ is closed under addition. Then, by Lemma 4, it follows that f is subadditive.

$\leftarrow$. Assume f is subadditive. Then by Lemma 4, epi(f) is closed under addition. Then by Lemma 3, epi(f) is convex. So then by Lemma 1, f is convex. $\square$

Finally, using Theorem 1, we are in a position to prove Wightman's proposition, that convexity implies and is implied by the second law of thermodynamics.

Proposition. Let $f : \mathbb{R}^n \to \mathbb{R}$ and assume it is homogeneous. Then, for all $x_1, x_2 \in \mathbb{R}^n$, $f(x_1 + x_2) = \inf \{f(x_1) + f(x_2)\}$ iff f is convex.

Proof. $\to$. Assume that $f(x_1 + x_2)$ is the infimum of $\{f(x_1) + f(x_2)\}$. It follows that $f(x_1+x_2) \leq f(x_1)+f(x_2)$ for all x_1, x_2, i.e. f is subadditive. It follows from Theorem 1 that f is convex.

$\leftarrow$. Assume that f is convex. Then by Theorem 1, f is subadditive. Let $A := \{f(x_1) + f(x_2)\colon$ for all $x_1, x_2 \in \mathbb{R}^n\}$. Let $a := \inf A$. We want to show that $a = f(x_1 + x_2)$. First we show that $f(x_1 + x_2) \in A$. It is the case that $f(\lambda(x_1 + x_2)) + f((1 - \lambda)(x_1 + x_2)) \in A$. From homogeneity, it follows that $\lambda f(x_1+x_2)+(1-\lambda)f(x_1+x_2) \in A$, i.e. $f(x_1+x_2) \in A$. From the fact that $f(x_1+x_2)$ is in A, it now follows that $f(x_1+x_2) = f(x_1')+f(x_2')$ for some positive x_1', x_2', and thus by the definition of the infimum, it follows that $f(x_1+x_2) \geq a$. Since a is the largest number such that $a \leq f(x_1) + f(x_2)$ and since $f(x_1 + x_2) \leq f(x_1) + f(x_2)$ for all $x_1, x_2 \in \mathbb{R}^n$, it follows that $a \geq f(x_1 + x_2)$. Therefore, we have that $a \leq f(x_1 + x_2)$ and $a \geq f(x_1 + x_2)$ so $a = f(x_1 + x_2)$. $\square$

This concludes our demonstration of Wightman's proposition, establishing a link between the second law and the convexity property of the fundamental relation.

2.3 Conclusion

I have presented the mathematical and conceptual foundations of geometric thermodynamics and set it in the context of other more well-known formulations of the discipline. It is one of the most mathematically rich formulations, allowing one to express the conventional first and second laws in a concise and precise manner. It also has welcome foundational implications for the oft-debated distinction between heat and work and the ‘anthropocentric’ view of thermodynamics.

Chapter 3

Representing Mixtures

3.1 Introduction

We all have an intuitive idea of what mixtures are; air is primarily a mixture of nitrogen and oxygen, and a martini is a mixture of gin and vermouth. But what sort of mathematical structure best describes a mixture? This question is an example of the well-known philosophical problem of determining what models best represent the phenomenon. In the case of mixtures, I identify two main candidates. The first, which I call the 'standard representation', is not a good representation of a mixture in spite of its widespread popularity. The second, which I call 'Gibbs' representation', is less widely adopted[1] but is, I argue, a much better representation. My argument will turn on a particular philosophical perspective concerning the representational capacities of mathematical structures. I will show that, once an adequate perspective on representation is adopted, Gibbs' representation leads to a better understanding of the notion of 'partial pressure' and of some associated

---

[1] A notable exception is Callen (1960) who follows Gibbs' presentation of thermodynamics as a whole.

thermodynamic results, Dalton's Law concerning the pressure of a mixture, and Gibbs' Theorem concerning the entropy of a mixture.

I begin in Section 3.2.1 by setting out two philosophical principles I will use to adjudicate between the two candidate representations. In Section 3.2.2 I will set out a mathematically precise formulation of thermodynamics, and a proposal for what it means to adequately represent physical systems as thermodynamic models. This will set the philosophical and formal background needed to compare the two candidate representations of a mixture. In Section 3.3, I argue that adopting the standard representation involves rejecting the two philosophical principles from Section 3.2.1. If we accept the principles and adopt Gibbs' representation, then two objections need to be addressed: (1) the ubiquitous use of the notion of 'partial pressure' in the context of mixtures, which appears to be absent in the Gibbs' representation, and (2) the absence of a representation of a mixing process based on Gibbs' representation. I address these objections in Sections 3.4 and 3.5.

3.2 Models of Thermodynamics

3.2.1 Adequate representations in thermodynamics

When physicists study physical systems, they represent those systems using the formal mathematical structures of some theory. For example, when we study the hydrogen atom, we study its representation as a quantum mechanical model; or when we study the motion of the planets in the solar system, we often study its representation as a model of classical gravitation. The same goes in the present case: when we study mixtures, we represent them using thermodynamic models.

We should therefore aim to formulate the concept of a thermodynamic model in a way that is appropriate for representing mixtures.

But what are thermodynamic models? Compare: a model of Hamiltonian mechanics is a symplectic manifold; a model of quantum mechanics is a Hilbert space with an algebra of observables; and a model of General Relativity is a Lorentzian manifold (M, g_{ab}). The fact that there are mathematically rigorous formal models available to represent these physical systems ensures that we have an adequate grip on what the structures of those physical systems are like. By contrast, it is striking that the thermodynamics literature reveals precious few mathematical definitions of a thermodynamic model. In the Chapter 2, I presented a mathematically precise definition of thermodynamic system, which brings thermodynamics on a par with the aforementioned areas of physics in this regard. In this Chapter, I will deploy this definition, and the formal machinery of the previous Chapter, to analyse mixtures.

But before we proceed, it is worth reflecting on why we should pursue precise mathematical definitions of our theoretical models in the first place. What physical or philosophical advantage does the existence of such formal models in classical, quantum or gravitation physics give us over the candidates for thermodynamic models described in Chapter 2?

The first advantage afforded to us by formal models is their usefulness to us as modellers: formal models allow us to say something about what an adequate model might look like for a particular physical system in some modelling context. By clearly laying out the formal structure the model uses to represent the system, we can see which aspects of the model represent particular aspects of the physical system. If a feature of a physical system is relevant in a particular context, then our model for it in that context had better be able to represent that physical feature. My point is better illustrated with the case I am concerned with in this paper: representing a mixture as a thermodynamic model. At the very least, this will mean that the thermodynamic model ought to represent all properties of the mixture deemed thermodynamically relevant. It follows that if the thermodynamic model does not represent some thermodynamically relevant property, then it is not an adequate thermodynamic representation of that system. More concisely, a model is not an adequate representation of a target if it fails to represent a property of the target it 'should' represent. Let us express this adequacy criterion more succinctly as follows:

> **Representational Adequacy (RA)**. Model X adequately represents target T in context C iff X possesses a property X_i which can be interpreted as representing some property T_i of T, deemed relevant in context C.

I do not which to commit to a more precise version of this criterion, since doing so would involve adopting a particular (possibly controversial) account of scientific representation. I anticipate that most perspectives on representation will deem this criterion to be reasonable, and that it can be formulated in whichever account of scientific representation one adopts[2].

As this principle is formulated, there is nothing to constrain what kind of thing the model X is; they may be concrete scale models just as well as mathematical structures. In my discussion, I am interested in the specific application of this

---

[2] An example of a precise account in which the principle might be formulated is that by Frigg and Nguyen (2020).

principle where X is a mathematical structure. Such structures have clearly identifiable properties X_i which can be used to represent relevant properties T_i of the target. Looking ahead, I will show that the standard representation of mixtures violates this adequacy criterion because it does not represent a feature of mixtures which is thermodynamically relevant, namely the volume of the mixture. Thus, if one accepts this criterion of representational adequacy, and one thinks that the volume of the mixture is thermodynamically relevant, then one ought to reject the standard representation.

The second advantage afforded to us by formal models is their usefulness to us as philosophers. It is our responsibility as philosophers to contemplate the interpretations of our mathematically formulated physical theories. Part of this investigation is to answer questions concerning (for example) the kind of mathematical models we take to represent physical systems, and the appropriate standard of equivalence of those models. Recent literature in general philosophy of science (such as the formal logic approach of Halvorson (2019)) and particularly in the context of the 'hole argument' of General Relativity (such as Weatherall (2018), Fletcher (2020) and Roberts (2020)), has illustrated how progress may be made on such questions if we first get clear on the mathematical structure lying at the foundations of a theory. Given the fruitfulness of this approach to philosophical problems in these areas, I would like to consider the extent to which such formal approaches may be fruitful in thermodynamics[3]. My view and argument in

---

[3] It should be noted the formal approaches to thermodynamics are not totally absent and have not gone unappreciated. Lieb and Yngvason (1999) is perhaps the most well-known work in formal thermodynamics and have made significant contributions to our understanding of its mathematical structure. However, they choose to follow in the axiomatic tradition of formalising physical theories. This is not the only approach we can take, since the geometric approach is also available. This work should therefore be seen as a step towards looking at the foundations of thermodynamics from a different angle in a way that complements other approaches.

this paper is that they can be. It is from this perspective that I approach the formalisation of thermodynamics in this Chapter.

One important step made in recent philosophical literature has been to draw attention to a principle concerning the capacity of mathematical models to represent physical situations. This principle was applied by Weatherall (2018) in the context of the hole argument:

> "[I]somorphic mathematical models in physics should be taken to have the same representational capacities. By this I mean that if a particular mathematical model may be used to represent a given physical situation, then any isomorphic model may be used to represent that situation equally well." (Weatherall, 2018, p. 332)

Further discussion of this principle is undertaken by Fletcher (2020), who expresses it as follows:

> **Representational Equivalence by Mathematical Equivalence (REME).** If two models of a physical theory are mathematically equivalent, then they have the same representational capacities. (Fletcher, 2020, p. 5)

Roughly speaking, the notion of the 'representational capacities' of a scientific model is taken to mean "the states of affairs that that model may be used to represent well" (Fletcher, 2020, p. 3). While this notion may seem intuitive, no precise definition of representational capacity exists, and its definition will surely depend on one's account of scientific representation. Fletcher avoids giving such a definition, on the basis that structural questions about, for example, how the mathematical equivalence of models constrains their representational capacities, are independent of one's definition of capacities. It will be similarly unnecessary for me to give a precise definition of representational capacities, since the example I am concerned with is stark enough that it should be accommodated by any definition.

I am not concerned with giving a defence of REME or related principles here.[4] Instead, I shall content myself with defending the following conditional claim: if one accepts REME, then one ought to reject the standard representation in favour of the Gibbs representation of mixtures.

In the following sections, I will set out what I take to be the formal structure of a thermodynamic model. I will then use this formalism to compare two candidate mixture representations against the same formal background. I will argue that the standard representation violates REME and the criterion for representational adequacy, and that we should therefore seek another representation.

3.2.2 Formal background

In this Section, I will apply the definition of thermodynamic system to the special case of the ideal gas. This provides a stepping stone to analysing the representation of mixtures of ideal gas, while also making connections to the more familiar ideal gas expressions from orthodox thermodynamics.

In this case, we assume the manifold has dimension $n = 4$ and coordinates that can be interpreted as S (the entropy), U (the internal energy), V (the volume) and N (the mole number). The fundamental relation in the entropy representation

---

[4] Fletcher defends the principle in the context of spacetime physics; it has been criticised by Belot (2018), Pooley (2021) and Roberts (2020), among others.

then takes the form

$$S = f(U, V, N), \tag{3.2.1}$$

and the homogeneity property is expressed as

$$\lambda S = \lambda f(U, V, N) = f(\lambda U, \lambda V, \lambda N) \tag{3.2.2}$$

for all values of U, V, N and for all positive values of λ.

To make contact with a more familiar expression of thermodynamics, we write the fundamental relation in differential form as

$$df = dS = \frac{\partial S}{\partial U}dU + \frac{\partial S}{\partial V}dV + \frac{\partial S}{\partial N}dN. \tag{3.2.3}$$

Now adopt the following definitions to denote the partial derivatives occurring in this equation:

$$\frac{\partial f}{\partial U} \equiv \frac{1}{T}; \qquad \frac{\partial f}{\partial V} \equiv \frac{p}{T}; \qquad \frac{\partial f}{\partial N} \equiv -\frac{\mu}{T} \tag{3.2.4}$$

where T, interpreted as the temperature, p, the pressure, and μ, the chemical potential, are known as the system's intensive parameters.[5] These intensive parameters are clearly functions of the extensive parameters U, V, N. Relationships

---

[5] The intensive parameters are often characterised as those which do not scale with the size of the system. This may be understood more precisely as follows. In the formal setup presented here, the intensive parameters are defined to be the partial derivatives of the fundamental relation with respect to the other extensive parameters. Since they are derivatives of one extensive parameter with respect to another, the intensive parameters are invariant under the scaling of the extensive configuration variables.

expressing the intensive parameters as functions of the extensive parameters are known as equations of state[6].

Writing the fundamental relation in differential form, we now arrive at what is commonly referred to as the *first law*,

$$dS = \frac{1}{T}dU + \frac{p}{T}dV - \frac{\mu}{T}dN. \tag{3.2.5}$$

It follows from Euler's theorem on homogeneous functions that[7]

$$S = \frac{1}{T}U + \frac{p}{T}V - \frac{\mu}{T}N. \tag{3.2.6}$$

This equation is known as the *Euler relation* and may be used to find the entropy of the ideal gas as a function of the extensive parameters U, V, N provided we know $1/T$, p/T and μ/T each as a function of U, V, N. From empirical investigations, we know the first two can be expressed in terms of a constant R, called the *ideal gas constant*, as

$$1/T = 3NR/2U; \qquad\qquad p/T = NR/V. \tag{3.2.7}$$

These are the familiar equations of state for the ideal gas. We can then work out an explicit expression for μ/T from Equations (3.2.7) using the Gibbs-Duhem

---

[6] It is yet another advantage of this formulation of thermodynamics that it puts the equations of state, familiar to all students of thermodynamics, in their proper conceptual place in the foundations. While important, they are actually derivative from the fundamental relation. An individual equation of state is not sufficient to capture all the thermodynamic information about a system. Adopting all the equations of state, however, is equivalent to adopting the fundamental relation itself, and hence is sufficient.

[7] For a proof of this, see Callen (1960, p. 47).

relation[8]. By a simple combination of these three equations, μ/T is found to be

$$\frac{\mu}{T} = R \ln \left[\frac{N^{5/2}}{U^{3/2}V} \right] + K, \tag{3.2.8}$$

where K is a constant. By substituting the expressions for the pressure, temperature and chemical potential into the Euler relation, we now derive the fundamental relation of the ideal gas to be

$$S = \left(\frac{3}{2} R \frac{N}{U} \right) U + \left(\frac{NR}{V} \right) V - \left(R \ln \left[\frac{N^{5/2}}{VU^{3/2}} \right] - K \right) N. \tag{3.2.9}$$

This equation may be simplified and substitutions of the equations of state may be made to express the entropy as a function of the other variables. This formulation has the merit of exhibiting the homogeneity of the fundamental relation, and emphasising the conceptual place of the equations of state as the derivatives of S with respect to U, V and N. Note that the fundamental relation's explicit form is given by the package consisting of the Euler relation and all the intensive parameters as a function of the extensive variables. In the rest of the paper, we will present the two candidate representations of mixtures like this.

We now turn to using this formulation of thermodynamics and the associated definition of thermodynamic model to discuss thermodynamic representations of gas mixtures.

---

[8] To derive the Gibbs-Duhem relation, we take the derivative of the Euler relation in Equation (3.2.6) and subtract from it Equation (3.2.3), the differential form of the first law. This yields $0 = Sd(1/T) + Vd(p/T) - Nd(\mu/T)$, which allows us to calculate the expression for μ/T using Equation (3.2.7).

3.3 Representing a Mixture in Thermodynamics

Intuitively, mixtures are made by combining two or more different substances in the very same volume without any chemical reaction happening. The challenge is to construct a mathematical representation of mixtures in thermodynamics so that we may model mixtures and their properties. In this Section, I will discuss and compare two candidates for such a representation: the 'standard representation' and 'Gibbs' representation'.

The standard representation is inspired by many discussions of thermodynamic mixtures, which all treat the 'partial pressures' of the gases as proper thermodynamic intensive variables, i.e. as partial derivatives of the fundamental relation with respect to volume variables[9]. The existence of partial pressures in the model is necessary for a popular representation of the mixing process based on semipermeable membranes: each membrane feels the partial pressure from a gas to which it is impermeable. This allows the gases to 'expand into each other' in the container. In contrast, Gibbs' representation is, as its namesake suggests, based on a description of mixture from Gibbs' 1878 work, which does not make explicit use of partial pressures as thermodynamic variables.

In the following Section, I will describe how our two representation principles require that we reject the standard representation in favour of Gibbs' representation. In short, my argument is this. If one thinks that the volume of the mixture is a relevant thermodynamic property which ought to be represented in the thermodynamic model, then, by RA, we should reject the standard representation in

---

[9] This representation is adopted (often implicitly, since many commentators do not use the formalism presented here) by Adkins (1983), Dieks (2018), Saunders (2018), van Kampen (1984), Planck (1903, §16).

favour of Gibbs'. On the other hand, one might think that the partial pressures ought to be represented in the thermodynamic model. In this case, by RA, we should reject Gibbs' representation in favour of the standard representation. However, if we follow this path, we violate REME and we may satisfy it by rejecting the standard representation in favour of Gibbs' representation. Therefore, if one accepts both RA and REME, we arrive at Gibbs' representation.

In Sections 3.4 and 3.5, details the consequences of accepting Gibbs' representation for Dalton's Law, Gibbs' Theorem and the mixing process. First though, we will examine each representation in detail.

3.3.1 The standard representation

I will argue that, if one accepts both of the adequacy principles for representation set out previously, then one ought to reject the standard representation.

For simplicity, let us consider an ideal gas mixture of two component gases, denoted 1 and 2. The standard representation of an ideal gas mixture is a thermodynamic system that I denote (M_s, f_s), where M_s can be given coordinates (U, V_1, V_2, N_1, N_2), and where the fundamental relation is

$$S = f_s(U, V_1, V_2, N_1, N_2) = \frac{1}{T}U + \frac{p_1}{T}V_1 + \frac{p_2}{T}V_2 - \frac{\mu_1}{T}N_1 - \frac{\mu_2}{T}N_2, \tag{3.3.1}$$

with

$$\frac{1}{T} = \frac{3}{2}R\frac{N_1 + N_2}{U}; \quad \frac{p_i}{T} = R\frac{N_i}{V_i}; \quad \frac{\mu_i}{T} = R\ln\left[\frac{N_i(N_1 + N_2)^{3/2}}{V_i U^{3/2}}\right] + K_i. \tag{3.3.2}$$

The model (M_s, f_s) is used to represent a mixture when the variables are given the following interpretations: S and U are the entropy and energy respectively; V_1 and V_2 are the volumes occupied respectively by gases 1 and 2; and N_1 and N_2 are the mole numbers of gases 1 and 2 in those volumes. However, with this interpretation, the model has an unsatisfactory feature: there are only coordinate variables for the individual gas volumes, and not for "the volume of the mixture" as a whole. Similarly, although the variables p_1 and p_2 are often interpreted[10] as the 'partial pressures' of the gases 1 and 2, there is no coordinate variable amongst the intensive variables of state space that represents the 'total pressure'. If one adopts the view that the volume and total pressure of a mixture are key physical features of the system that merit representation in the model, then this model fails to do it. Hence, (M_s, f_s) violates the Representational Adequacy criterion and we ought to seek a model which does account for this key physical feature.

One response to this argument is to claim that (M_s, f_s) can represent the volume of the mixture so long as we stipulate that V_1 and V_2 have the same value, say V, and that they 'overlap', or refer to the very same place in physical space. Let me refer to this assumption as *Volume Overlap.* This view is often implicit in discussions of mixtures when the concepts 'volume of gas 1' or 'volume of gas 2' are used in the context of describing mixtures as thermodynamic systems.

However, there is a problem with this view. Due to the interpretation of V_1 and V_2 as 'the volume of the individual gases', it is perfectly possible to use (M_s, f_s) to represent the physical system consisting of two spatially separate, independent volumes V_1 and V_2. As soon as one chooses not to model the volume of the mix-

---

[10] Some examples of authors who seem to adopt partial pressures as genuine thermodynamic variables of a mixture, and hence would adopt this standard representation of a mixture are Planck (1903, pp. 11–12), van Kampen (1984) and Adkins (1983, p. 217).

ture, but rather the volumes of the individual gases, there is nothing about the mathematical model which makes it a representation of a mixture as opposed to a representation of a system with two spatially separate volumes. The formalism does allow us to assert that volumes V_1 and V_2 have the same value, but it does not have the capacity to represent *Volume Overlap*. What makes (M_s, f_s) a representation of a mixture as opposed to a representation of a physical system of two separate volumes is not any mathematical structure in the formal model, but rather the informal *Volume Overlap* assumption.

In sum, the model of a mixture is actually different from the model of two separate volumes; the model which represents the mixture is '(M_s, f_s) + *Volume Overlap*' and the model which represents the two separate volumes is just (M_s, f_s). But, if we are limited to the representational capacities of the thermodynamic formalism, the models are one and the same. We may understand the problem of representing mixtures as (M_s, f_s) + *Volume Overlap* in a different light as a violation of REME.

If we let $M_1 = (M_s, f_s)$ + *Volume Overlap* and $M_2 = (M_s, f_s)$ then we have the situation where M_1 has the capacity to represent mixtures but M_2 does not. They are mathematically equivalent (in fact, they are identical) in thermodynamics, since the volume overlap assumption is not manifested formally in the model, but they have different representational capacities, in clear violation of REME. It follows that using M_1 violates this criterion for formal representation. Since M_1 and M_2 have different representational capacities, they had better not be mathematically equivalent.

To fix this problem and abide by REME, we can implement *Volume Overlap* formally as part of the mathematical model by simply having one volume variable,

V. The ambiguity about where in space the gases are located is removed, since there is only one place they can be, and there is no need to stipulate that the volumes of the gases both have the same value, since there is only one volume variable 'containing' the mole numbers N_1 and N_2. This means we can now model something which we could not under (M_s, f_s): we can model the key physical feature, 'the volume of the mixture', as V. Thinking this way leads us on to the next candidate for the representation of a mixture: Gibbs' representation.

3.3.2 Gibbs' representation

The alternative representation of a mixture I would like to advocate adheres to REME, and is able to represent the volume of the mixture directly. One might think of this as a formal implementation of *Volume Overlap*: we represent an ideal gas mixture of k components as the system (M_G, f_G) where M_G may be given coordinates $(U, V, N_1, \ldots, N_k)$ and the fundamental relation is

$$S = f_G(U, V, N_i, \ldots, N_k) = \frac{1}{T}U + \frac{p}{T}V - \sum_{i}^{k} \frac{\mu_i}{T} N_i \tag{3.3.3}$$

where

$$\frac{1}{T} = \frac{3}{2}R\frac{(\sum_i^k N_i)}{U}; \quad \frac{p}{T} = R\frac{(\sum_i^k N_i)}{V}; \quad \frac{\mu_i}{T} = R\ln\left[\frac{N_i\left(\sum_i^k N_i\right)^{3/2}}{VU^{3/2}}\right] + K_i. \tag{3.3.4}$$

On this characterisation, we must think of an ideal gas mixture as some number of different substances with mole numbers N_i, all with the same temperature, occupying the very same volume V. This definition incorporates the previously informal *Volume Overlap* as explicit formal structure in the model, by simply having only one volume variable which we may easily interpret as the 'volume of the mixture'.

This representation not only has intuitive appeal, but also historical precedent: it is the same as that adopted by Gibbs. While Gibbs did not yet have the formal tools to adopt the definition of a thermodynamic system I propose, it is clear that he takes (M_G, f_G) to be the definition of an ideal gas mixture. That is, he takes the extensive variables which describe the thermodynamic state of a mixture to be $S, U, V, N_1, ..., N_k$ for a mixture of k components:

> "[I]f we consider the matter in the mass as variable, and write $N_1, ..., N_k$ for the quantities of the various substances $1, ..., k$ of which the mass is composed, U will evidently be a function of S, V, $N_1, ..., N_k$ and we shall have for the complete value of the differential of U
>
> $$dU = TdS - pdV + \mu_1 dN_1 + ... + \mu_k dN_k$$
>
> $\mu_1, ..., \mu_k$ denoting the differential coefficients of U taken with respect to $N_1, ..., N_k$." (Gibbs, 1878, p. 116)

Thus, Gibbs takes the variables for a thermodynamic system of a homogeneous gas consisting of various substances of mole number N_i to be described by the extensive variables, $S, U, V, N_1, ..., N_k$. This differs from the variables for a pure substance only in the number of variables N_i denoting the quantities of the substances.

Having settled on his choice of variables, Gibbs begins his derivation of the fundamental relation of an ideal gas mixture, by assuming the following principle:

> "The pressure in a mixture of different gases is equal to the sum of the pressures of the different gases as existing each by itself at the same temperature and with the same value of its [chemical] potential." (Gibbs, 1878, p. 215)

From this principle[11], Gibbs is able to deduce that the pressure of an ideal mixture of n components is,

$$p = \sum_i^n C_i e^{\mu_i/T} T^{5/2}. \tag{3.3.7}$$

That is, the pressure of the mixture is the sum of the individual pressures each of the gases would have if they were at temperature T and chemical potential μ_i.

Gibbs proposes that Equation (3.3.7) can be viewed as the "fundamental equation" describing a mixture:

> "It will be legitimate to assume this equation provisionally as the fundamental equation defining the ideal gas mixture, and afterwards to justify the suitableness of such a definition by the properties which may be deduced from it." (Gibbs, 1878, p. 216)

As it is currently expressed, it is not a relation between the extensive variables and so does not fit the description of a fundamental relation in the sense it

---

[11] To understand how to express this principle in the formalism, note that, for a pure ideal gas, μ/T as a function of pressure and temperature is

$$\frac{\mu}{T} = R\ln p - \frac{5}{2}R\ln T + K \tag{3.3.5}$$

where K is a constant. We may rearrange this for p to give

$$p = Ce^{\mu/T}T^{5/2} \tag{3.3.6}$$

where C is a constant absorbing all the constant terms that appear in the course of the substitutions. This equation allows us to express the principle Gibbs described as Equation (3.3.7).

was introduced in Section 3.2. But it is possible to show that Equation (3.3.7) is a fundamental relation, since it is the full Legendre transform of the fundamental relation between the extensive variables. To extract the explicit form of the fundamental relation f_G, we first note that Equation (3.3.7) implies[12]

$$p = \sum_i^k \frac{N_i T}{V}; \tag{3.3.8}$$

$$S = \sum_i^k \left(\frac{3}{2} N_i R \ln T - N_i R \ln \frac{N_i}{V} + N_i C_i \right). \tag{3.3.9}$$

We now substitute the expressions for S/V, N_i/V and p into the Euler relation for the ideal gas mixture (Equation (4.3.1)), which yields $U = (3/2)R\left(\sum_i^k N_i\right)T$. Taking these three equations together, we have worked out all of the equations of state of the gas mixture:

$$\frac{1}{T} = \sum_i^k \frac{3}{2} R \frac{N_i}{U}; \quad \frac{p}{T} = \sum_i^k R \frac{N_i}{V}; \quad \frac{\mu_i}{T} = R \ln \left[\frac{N_i \left(\sum_j^k N_j \right)}{V U^{3/2}} \right] + K_i. \tag{3.3.10}$$

To see the fundamental relation (as a relation between the extensive variables) of the ideal gas mixture, f_G, in its explicit form, we may substitute the equations of state into Equation (4.3.1).

Thus, the thermodynamic representation of an ideal gas mixture may be derived from an empirical principle, and how all of its familiar thermodynamic properties are derived. This also concludes my argument that Gibbs' representation adheres to both principles laid out in the last section and that the standard rep-

---

[12] I show this, as well as the fact that Equation (3.3.7) is the full Legendre transform of $f_G(U, V, N_1, \ldots, N_k)$, in detail in Appendix A.2.

resentation violates them both.

I am now going to consider two objections to my line of reasoning. The first appeals to the apparent ubiquitous use of the notion of 'partial pressure' in discussion of mixtures and, in particular, to the statement of certain laws concerning mixtures: Dalton's Law about the pressure of a mixture and Gibbs' Theorem about the entropy of a mixture. I address this in Section 3.4 by demonstrating that both laws are in fact consequences of Gibbs' representation where no notion of partial pressure appears. The second objection stems from the fact that a typical and popular representation of mixing based on semipermeable membranes depends on the notion of partial pressure and so Gibbs' representation (without such a notion), cannot be the basis for a representation of mixing. I address this by providing a representation of mixing based on Gibbs' representation of mixture.

3.4 Dalton's Law and Gibbs' Theorem

An attractive feature of Gibbs' representation of a mixture is that there is only one volume variable V, and hence only one pressure p. This may prompt an objection from those who see partial pressures and partial volumes as essential variables for characterising a mixture: how can we describe mixtures without the concept of partial pressure? To see partial pressures and mixtures intimately associated with each other is forgivable: Dalton's Law for mixtures is sometimes stated in terms of partial pressures,[13] and Gibbs' Theorem for mixtures expresses an analogous principle with respect to entropy.

In fact, neither Dalton's Law nor Gibbs' Theorem depend on the concept of

---

[13] For example, see Adkins (1983, p. 214) and Blundell and Blundell (2010, p. 60).

partial pressure. Indeed, Equations (3.3.8) and (3.3.9), which are consequences of Gibbs' representation, are statements of Dalton's Law and Gibbs' Theorem respectively: they are both mathematical consequences of a fundamental relation which does not have partial pressure as an intensive variable. The purpose of this Section, then, is to show how something that *looks like* partial pressures may appear, and to go on to explain how they should properly be interpreted. We begin with an examination of the mathematical origin of the laws.

Dalton's Law is the principle that "the pressure in a gas-mixture is equal to the sum of the pressures which the component gases would possess if existing separately with the same volume at the same temperature." (Gibbs, 1878, p. 217). Gibbs' theorem states an entirely analogous principle with respect to entropy, which Adkins (1983, p. 217) takes to be the fact that "the entropy of the mixture is the sum of the entropies that the component gases would have if present alone", and Callen (1960, p. 335) takes Gibbs' theorem to be the fact that "the entropy of a mixture of ideal gases is equal to the sum of the entropies that each of the gases would have if it alone were to occupy the same volume at the same temperature."

Let us understand in more depth why these propositions are true. We will just cover the case of Dalton's Law, since a similar argument applies to entropy and Gibbs' Theorem. The pressure of a single pure ideal gas i as a function of volume and temperature is given by $p_i = N_i T_i / V_i$. If we evaluate the pressures of the pure ideal gases at temperature value $\widetilde{T}$ so that $T_i = \widetilde{T}$ for all i and at volume value $\widetilde{V}$ so that $V_i = \widetilde{V}$ for all i, then the pressure of each gas is given by $p_i = N_i \widetilde{T} / \widetilde{V}$. The pressure of a mixture evaluated at $V = \widetilde{V}$ and $T = \widetilde{T}$ is given by Equation (3.3.8) as $p = \sum_i N_i \widetilde{T} / \widetilde{V}$. By comparing the values of the pressure of the pure ideal gases and the pressure of the mixture, it is clear that the pressure of the mixture evaluated at temperature and volume values $\widetilde{T}$ and $\widetilde{V}$ is equal to the sum of the pressures of the individual pure gases evaluated at the same temperature and pressure values. This coincidence of values of the pressure of a mixture and the sum of the pressures of individual gases is Dalton's Law.

My rather pedestrian explication of Dalton's Law serves to highlight that the notion of 'partial pressure' plays absolutely no role in Dalton's Law. The only thermodynamic pressure variables playing a role in this argument are the individual pressures of the separate ideal gases, and the pressure of the mixture. Now, we could choose to define $\widetilde{p}_i := N_i\widetilde{T}/\widetilde{V}$ and call it the 'partial pressure' so that the pressure of the mixture may now be written $p = \sum_i \widetilde{p}_i$, i.e. the sum of the partial pressures. But this definition, at best, serves as nothing more than a mathematical fiction,[14] and at worst as method of deceiving us into thinking that partial pressures are genuine thermodynamic variables. My point can be summarised more succinctly: partial pressures are not thermodynamic variables associated with any thermodynamic model featured in the argument establishing Dalton's Law; as a result, they lie strictly outside the representational capacity of that model. Partial pressures are at best extra-theoretic constructs that illustrate a coincidence in the values of pressures of individual gases and a of mixture. It seems rather ironic that Dalton's Law, which seemingly makes unavoidable reference to the notion of partial pressure, is a derivable property of the thermodynamic system (M_G, f_G), whose main difference from (M_s, f_s) is that it does not have partial pressures as thermodynamic variables!

---

[14] Callen (1960, p. 337) writes "The partial pressures are purely mathematical constructs with no direct physical meaning."

3.5 Mixing and Equilibration

A second objection to the Gibbs representation might go as follows: this representation of a mixture cannot be right, because it does not allow one to say how a mixture is the result of mixing. How could this be possible without partial pressures and the standard representation? If we are to follow Gibbs in adopting the (M_G, f_G) representation and giving up up the 'partial pressures' as thermodynamic variables, we have to give up a popular and typical representation of the process of mixing based on (M_s, f_s) as each gas pushing on a semipermeable membrane with its corresponding partial pressure.[15] In this Section, I will show how this representation is capable of representing mixing as equilibration in a conceptually clear way, and indeed one that is more precise than many other accounts.

3.5.1 Gibbs' representation of mixing

Mixing is typically described as the process where both samples of gas, each initially confined to their half of a container, finally come to occupy the full volume of the container when the partition dividing them is removed. Thus, Dieks writes that mixing is the gases "expanding into the same final volume" (Dieks, 2018) and van Lith writes that we have a mixture when the gases are "spread out over the whole container" (van Lith, 2018). These suffice for intuitive grasps on the concept but the challenge now is to represent the process of mixing in the thermodynamic formalism. I will take the uncontroversial stance that the mixing process is an instance of equilibration: a thermodynamic system coming to equilibrium.

It is readily admitted that equilibrium thermodynamics has nothing to say

---

[15] The way of describing the mixing process is very popular and is described by many textbooks and commentators. It goes back, at least, to Planck (1903).

about the process of mixing, because it is obviously a non-equilibrium process. But it is possible to give an account of equilibration in equilibrium thermodynamics, as I have shown in Chapter 2.

Let us briefly illustrate how these definitions work by considering an ideal gas decomposed into two subsystems along the lines outlined in the definition of K-equilibirum. For the composite system to be in equilibrium, the values of the variables $U^{(1)}, V^{(1)}, N^{(1)}$ and $U^{(2)}, V^{(2)}, N^{(2)}$ must be such that they extremise the entropy of the composite system S which is given by $S = S^{(1)} + S^{(2)}$. Thus we have that

$$S = S^{(1)} + S^{(2)} = \frac{1}{T^{(1)}}U^{(1)} + \frac{p^{(1)}}{T^{(1)}}V^{(1)} - \frac{\mu^{(1)}}{T^{(1)}}N^{(1)} + \frac{1}{T^{(2)}}U^{(2)} + \frac{p^{(2)}}{T^{(2)}}V^{(2)} - \frac{\mu^{(2)}}{T^{(2)}}N^{(2)}. \tag{3.5.1}$$

To find the extremum of S we differentiate and set $dS = 0$. If we allow variation in all the extensive variables, then we have, for all $X = U, V, N$, $X^{(1)} + X^{(2)} = X$ where X is constant implying $dX^{(2)} = -dX^{(1)}$. Therefore we find:

$$0 = \left(\frac{1}{T^{(1)}} - \frac{1}{T^{(2)}}\right) dU^{(1)} + \left(\frac{p^{(1)}}{T^{(1)}} - \frac{p^{(2)}}{T^{(2)}}\right) dV^{(1)} - \left(\frac{\mu^{(1)}}{T^{(1)}} - \frac{\mu^{(2)}}{T^{(2)}}\right) dN^{(1)}. \tag{3.5.2}$$

dS must vanish for arbitrary values of $dU^{(1)}$, $dV^{(1)}$ and $dN^{(1)}$ so we find the conditions for thermodynamic equilibrium to be

$$\frac{1}{T^{(1)}} = \frac{1}{T^{(2)}}; \qquad \frac{p^{(1)}}{T^{(1)}} = \frac{p^{(2)}}{T^{(2)}}; \qquad \frac{\mu^{(1)}}{T^{(1)}} = \frac{\mu^{(2)}}{T^{(2)}}. \tag{3.5.3}$$

Therefore, the ideal gas is at thermodynamic equilibrium when the subsystems have equal temperature, pressure and chemical potential. These conditions allow us to calculate the equilibrium values of the subsystems' extensive variables. For example, substituting the equation of state $U = (3/2)NRT$ into $1/T^{(1)} = 1/T^{(2)}$ yields $U^{(1)}/N^{(1)} = U^{(2)}/N^{(2)}$. Since $U = U^{(1)} + U^{(2)}$, it follows after some rearrangement that $U^{(1)}/N^{(1)} = U^{(2)}/N^{(2)} = U/N$.

Having understood the definition and structure of equilibrium, we are now able to bring what we have learnt to bear on mixing and in particular how to understand mixing based on the thermodynamic system (M_G, f_G).

3.5.2 Mixing based on Gibbs' representation

The alternative view I will present is the following: mixing is the transition from the unmixed state (where the values of the subsystems' extensive variables do not extremise the entropy) to the mixed state (where they do) as a result of the removal of the constraint on the flow of the mole number of each gas between the subsystems.

It is assumed that the temperatures and pressures on each side of the partition are equal. This means that the system is already in mechanical and thermal equilibrium, leaving only the mole number constraints to be removed. In the context of (M_G, f_G), the only equilibrium there is left to achieve is chemical equilibrium: removing the partition between the subsystems will allow the flow of the mole number quantity of each type of gas between the subsystems until the chemical potentials of each subsystem become equal.

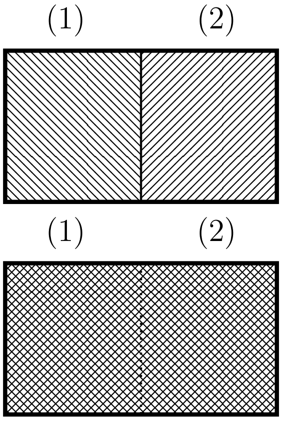


Figure 3.1: Depiction of the mixing process according to the Gibbs representation. The initial, unmixed system is depicted on top and the final, mixed system on the bottom.

Variable	Initial	Final
$U^{(1)}$	$U/2$	$U/2$
$U^{(2)}$	$U/2$	$U/2$
$V^{(1)}$	$V/2$	$V/2$
$V^{(2)}$	$V/2$	$V/2$
$N_1^{(1)}$	N	$N/2$
$N_1^{(2)}$	0	$N/2$
$N_2^{(1)}$	0	$N/2$
$N_2^{(2)}$	N	$N/2$

Table 3.1: Table illustrating the initial and final values of the extensive variables of the subsystems, before and after the removal of the constraint on the mole number.

The final equilibrium values of the mole number quantities are calculated by extremising the entropy.

$$0 = dS = \left(\frac{\mu_1^{(1)}}{T} - \frac{\mu_1^{(2)}}{T}\right) dN_1^{(1)} + \left(\frac{\mu_2^{(1)}}{T} - \frac{\mu_2^{(2)}}{T}\right) dN_2^{(1)}, \qquad (3.5.4)$$

concluding that the condition for equilibrium is $\mu_1^{(1)}/T = \mu_1^{(2)}/T$ and $\mu_2^{(1)}/T = \mu_2^{(2)}/T$. Imposing this condition implies that the gases expand to fill the volume of the container, with final values as illustrated in Table A.2.

This representation of mixing is rather different[16] compared with the representation based on (M_s, f_s) and its associated semipermeable membranes. In the

---

[16] Nevertheless, the entropy increase due to mixing is the same in each case, as I show in Appendix A.3.

process based on (M_s, f_s), mixing is initiated by allowing the 'partial pressures' of each gas to push on each semipermeable membrane, expanding the volume which each gas occupies. In the process based on (M_G, f_G), the pressures and temperatures of each subsystem are assumed to be equal, so mixing is initiated by removing the mole number constraint.

In the face of the problems described in Section 3.3.1, and given that one can recover the standard numerical answer to the entropy increase on mixing on the basis of Gibbs' representation of a mixture, (M_G, f_G), it seems reasonable to abandon the standard representation (M_s, f_s) + *Volume Overlap* as a representation of mixture and instead adopt (M_G, f_G) as the better representation.

3.6 Conclusion

This Chapter has addressed two issues. The first was justifying the use of mathematical models in thermodynamics. The need for such a formal representation stems from the fact that physicists study physical systems in the context of some formal theory and it is therefore necessary to know what formal structure is representing the target physical system.

The second issue addressed was the existence of two incompatible representations of mixtures in the thermodynamic literature. I used the definition of a model in thermodynamics to make these two representations precise and compare them against the same formal background. I argued that the representation based on Gibbs' description of mixtures was preferable to the standard representation which takes partial pressures seriously. Abandoning the standard representation meant we also had to abandon the standard and popular representation of the mixing process based on semipermeable membranes which made explicit use of partial pressures. I presented an alternative representation mixing based on Gibbs' representation of mixture and clarified the role of partial pressures in thermodynamics in the context of Dalton's Law.

Chapter 4

Homogeneity and Identity in Thermodynamics

4.1 Introduction

The concepts used in stating or deriving propositions of a formal theory are only rigorously understood when they are defined in terms of the formalism of that theory. To illustrate, consider this example: "Let (M, g_{ab}) be a temporally orientable spacetime" may be an assumption to prove some theorem in general relativity. In this proposition, both 'spacetime' and 'temporally orientable' are concepts which are given mathematically precise definitions in the formalism of general relativity. Without such definitions, the proposition wouldn't have quantitative meaning in the theory, even if we may have intuitive grasps on what it means to be temporally orientable. Although much of physics seems to naturally respect this practice, some of thermodynamics does not.

I will argue that 'identity' is a relation used without any formal representation in thermodynamics, and yet is still used in stating or deriving thermodynamic propositions. For example, a statement such as "Let A and B be two identical gases" cannot be expressed in the thermodynamic formalism because, I claim, there is currently no agreed formal representation of the identity relation in the theory. The relation is given only qualitative meaning drawn from extra-theoretic considerations by, for example, appealing to properties of the relata which the theory does not have the capacity to represent. There is nothing inherently wrong with doing this; physicists often proceed with reasoning without fully defining certain concepts, relying on intuitive or extra-theoretic understanding of the terms. This is often necessary when new physics is being formulated and new paths of discovery being forged.

However, it is important that they should be given clear thermodynamic meaning in applications where certain quantitative judgments are needed, and is also important for philosophers of physics who are interested in understanding the conceptual foundations of a theory in clear and rigorous terms. The dangers of not precisely defining concepts in theoretical reasoning are illustrated in this particular case by the existence of a purported paradox in deriving the entropy of mixing: Gibbs' paradox. Roughly speaking, (one version of) the paradox is that we can apparently derive a non-zero entropy of mixing when the gases are identical, contradicting the view that there should be no entropy of mixing in this case. If we are to establish a link between identity and no mixing entropy in thermodynamics then we had better be able to formulate a deductively valid argument in the thermodynamic formalism beginning with a thermodynamic proposition which we interpret as identity and ending with a thermodynamic proposition stating that there is no mixing entropy.

This is not to say that representation of the identity relation has eluded attention entirely. Both physicists and philosophers who comment on the paradox mostly do take care to define it; in fact, most see identity's representation (rightly, in my view) as the heart of the problem and the key to the solution. Rather, my point is that existing representations do not respect the good practice described above: they are not thermodynamic, but extra-theoretic, appealing to concepts and ideas which thermodynamics does not have the capacity to represent. This is often heuristically useful and helps us muddle through abstract reasoning by latching on to more familiar situations. But I am interested in those investigations where taking care in this regard is to avoid paradoxical conclusions. Gibbs' paradox is a case in point.

I will address this issue by proposing a representation of identity using the geometric formulation of thermodynamics set out in Chapter 2. This leads to a new perspective on the philosophical concept of identity and the conceptual and formal structure of thermodynamics that is free from Gibbs' paradox. The basic idea I will present in detail in the following sections is this. I first use the geometric formulation to highlight the idea that thermodynamic systems are homogeneous, that is, they 'look the same' on all scales. I then use the geometric formulation to characterise a mixture as a composite system whose components are the subsystems, looking to define a relation between the subsystems which represents (non-)identity. The criterion I then propose defines subsystems of a composite as identical when they each scale to the composite. The components of a mixture are different in the sense that they do not scale to the composite. In this way, the representation of the identity relation makes essential use of the homogeneity assumption.

In Section 4.2, I introduce Gibbs' paradox in thermodynamics and explain how it is driven by a lack of a clear understanding of what it means for thermodynamic models to be 'identical'. Sections 4.2.1 and 4.2.2 describe in some detail an early and prominent response to the paradox by van Kampen, seen by many[1] as the authoritative treatment of Gibbs' paradox in thermodynamics, before noting what I see a key difficulty with this response in Section 4.2.3: the lack of a clear and precise criterion of identity for thermodynamic models. The rest of the paper is geared towards providing this criterion. Section 4.3 provides an analysis of "mixture" in terms of the composite system definition of Chapter 2. Finally, in Section 4.4, I give a thermodynamic expression for the identity of thermodynamic models in the geometric formalism that is based on the assumption of homogeneity, and I explain how this solves Gibbs' paradox.

4.2 Gibbs' Paradox in Thermodynamics

Let two samples of gas at equal temperatures (and equal mole number and volume for simplicity) be separated by a partition. When the partition is removed, do the gases mix? What happens to the entropy of the system: does it increase or stay the same? Standard answers to these questions say that it depends whether the gases are 'identical' or 'different': the gases mix and there is an increase in entropy if the gases are different, known as the 'entropy of mixing'. But they are identical when the samples of gas do not mix and there is no entropy increase. The difference between the two situations is depicted in Figure 4.1.

---

[1] See, for example, Saunders (2006), Saunders (2013), and Saunders (2018) whose treatment of the Gibbs paradox in thermodynamics follows van Kampen, van Lith (2018, p. 7) praises van Kampen's analysis and reviews of thermal physics (e.g. Uffink (2007) and Frigg (2008)) cite van

(a) Two different samples of gas separated by a partition.

(b) Two identical samples of gas separated by a partition.

(c) The mixture is formed when the partition is removed.

(d) Nothing happens when the partition is removed.

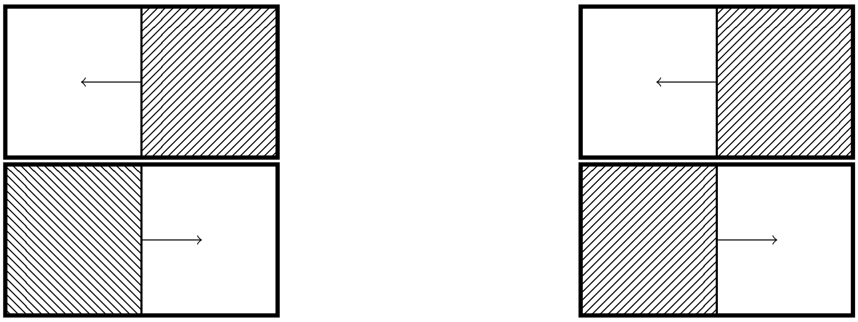
Figure 4.1: The Gibbs paradox setup. Intuition dictates that nothing happens to the system when the partition is lifted between two identical samples of gas, while mixing happens when the partition is lifted between different samples of gas.

Confusions start to arise when we inquire into the precise nature of this identity or difference and how these notions are represented in the thermodynamic formalism. Versions of the paradox exist in both thermodynamics and statistical mechanics. Here we are interested in the thermodynamic version. The statements of the thermodynamic paradox that exist in the literature fall into two categories. The first:

GP1 On one popular definition, the entropy of mixing does not appear to have any formal component whose value depends on whether the two gases are distinct, thus it seems as though there would still be a non-zero entropy of mixing when the partition is lifted between two identical samples of ideal gas, contradicting the intuition that there is no entropy of mixing for identical gases.([2])

Kampen as an authoritative treatment of Gibbs' paradox.

[2] Bridgman (1961, p. 1698), van Kampen (1984, p. 306), Elwell and Pointon (1972, p. 145),



(a) (b)

Figure 4.2: Lifting the partition between two samples of gas on each side of a partition in a container is often modelled as each half of gas undergoing a free expansion in separate containers. Modelling it this way does yield the correct value for the entropy of mixing different gases (Figure 4.2a), but also results in the same entropy of mixing for identical samples of gas (Figure 4.2b).

GP1 is driven by two conflicting representations of happens when the partition is lifted. On one representation, the lifting of the partition allows the samples of gas on each side to expand into the other half of the container, inviting us to imagine mixing as two spatially separate expansions occurring independently (see Figure 4.2).

But if we represent mixing in this way, then the result of lifting the partition is completely independent of whether the gases are identical or different, since it is explicitly represented as two spatially separated samples expanding in their own containers; the identity or difference of the substances has no bearing on the result of such independent expansions. This contradicts the result of another representation of mixing depicted in Figure 4.1 which indicates that the result of removing the partition does depend strongly on the identity of the gases; mixing happens when the partition is lifted between two different samples of gas, while

---

Adkins (1983, p. 217), Fong (1963, pp. 48–49), Partington (1913, p. 274), Fay (1965, p. 218), Pierce (1968, p. 319), Muller (1985, p. 193).

there is no mixing if the gases are identical.

The second version of the paradox is the following:

GP2 Given that there is no entropy change when the samples of gas are identical, then the entropy change appears to vary discontinuously from a non-zero value to zero as the gases go from being different to identical. This contradicts a principle, generally held dear by the physics community, that there are no discontinuities[3] in physical quantities.([4])

GP2 is driven by the intuition that some substances are more similar than others, and that this similarity can be imagined as varying continuously. This similarity and continuous variation, it is argued, should be manifest in a continuous variation of the mixing entropy as the substances go from being different to identical. Commentators who are worried by this version of the paradox aim to remove the discontinuity by demonstrating that the entropy does vary continuously as a function of some variable interpreted as the 'degree of similarity' between the substances. Those who are not worried by this version of the paradox attempt to explain why this discontinuity is not a target of this principle.

These two versions of the paradox identify very different problems. Those concerned with GP2 assume that the mixing entropy for identical gases is zero, seeing the problem to be that the mixing entropy does not vary continuously down to zero as the gases become more similar. By contrast, those concerned by GP1 see

---

[3] This principle is seen to originate historically with Leibniz (*Natura non facit saltum*: 'nature does not jump').

[4] Denbigh and Redhead (1989, p. 284), Dieks (2014, p. 1305), Dieks (2011, p. 368), Barsky (2017, p. vii), Denbigh and Denbigh (1985, p. 76), Kondepudi and Prigogine (1998, p. 157), Landsberg (1978, p. 74), Roberts and Miller (1960, p. 379), Sears and Salinger (1975, p. 241), Smith (1952, p. 137), Zemansky and Dittman (1997, p. 391), Swendsen (2018), Dieks (2018), van Lith (2018), Uffink (2007, p. 1000), Klein (1958, p. 80).

the problem precisely in the assumption that GP2 takes for granted: the mixing entropy for identical gases being zero. Since GP1 identifies an issue in something GP2 takes for granted, it is a deeper problem and must be solved first[5]. We will therefore focus our efforts in this paper to solving and understanding GP1.

Needless to say that solutions to GP1 abound and, furthermore, commentators on the paradox agree on the problem and solution. In fact, commentators on Gibbs' paradox universally agree that the thermodynamic paradox has been solved, seeing the interesting and difficult work as articulating and solving the statistical mechanical paradox. My main contention in this paper is that statements of and solutions to the thermodynamic Gibbs paradox suffer from conceptual and mathematical problems, which arise from using a qualitative, extra-theoretic definition of identity. Once we clarify this, we end up with a very different, but much more secure, solution.

The strategy typically taken to solve GP1 is to argue somehow that the representation of mixing as two free expansions cannot apply when the gases are identical. In the next Section, I will showcase an extremely popular and oft-cited presentation of this line of reasoning. I will then critically analyse this account and identify mathematical and conceptual problems with it which I will go on to solve.

---

[5] Nevertheless, I will offer some brief comments on GP2. In short, I agree with Dieks (2011) that the discontinuity is neither surprising nor problematic once we have established that the entropy of 'mixing' identical gases is zero, since the discontinuity is not in the physical quantity but is rather a 'conceptual' discontinuity in going from considering identical gases to non-identical gases.

4.2.1 Van Kampen's Account

Van Kampen begins by deriving the entropy as a function $S(p,T)$ of pressure p and temperature T. He leaves the derivation of this function, very typical in standard textbooks on thermodynamics, to the reader, but it is instructive for our purposes to spell it out in detail[6].

We begin by stating what is known as the First Law of thermodynamics for a thermodynamic model consisting of variables for energy U, temperature T, entropy S, pressure p, and volume V:

$$dU = TdS - pdV. \tag{4.2.1}$$

Now let N be the (constant) mole number and R the ideal gas constant. Adopting the equations of state $U = (3/2)NRT$ and $pV = NRT$ for the ideal gas, and observing that $dU = (3/2)NRdT$, we find that, after some rearrangement, Equation (4.2.1) becomes,

$$dS = \frac{3NR}{2T}dT + \frac{NR}{V}dV. \tag{4.2.2}$$

Integrating, we get $S(V,T) = \frac{3}{2}NR\ln T + NR\ln V + K$, and if we substitute $V = NRT/p$, we get van Kampen's expression for $S(p,T)$:

$$S(p,T) = \frac{5}{2}NR\ln T - NR\ln p + C \tag{4.2.3}$$

---

[6] The derivation presented in the main text dodges a problem stemming from an inconsistency in van Kampen's presentation, namely: is the number of molecules N a constant or a variable? As I explain in Appendix A.1.1, he seems to treat it as both. I argue that his expression for $S(p,T)$ can be derived rigorously assuming that N is a variable throughout.

where C is a constant of integration absorbing terms containing N and R.

At this point, we depart typical entropy derivations and follow van Kampen's unique argument to calculate the entropy on mixing identical gases and different gases. Van Kampen points out that C cannot be a function of p or T, but may be a function of anything else including N. He then supplies an argument for determining the dependence of C on N (and hence entropy's dependence on N) on the basis of the following two propositions:

> "**Convention 1**. If I have a second system, identical with the first one, I agree to assign to it the same C and therefore the same $S(p, T)$. This is obviously convenient but not logically necessary and by itself contains no information about the physical world.
>
> **Convention 2**. Two separate systems may be regarded as forming together a single system. I agree to take as the entropy of the combined system the sum of the entropies of the two parts. Of course, each part must be in equilibrium but they don't need to be in equilibrium with one another. Hence this convention may serve to define the entropy of systems that are not in equilibrium themselves but can be divided into subsystems that are." (van Kampen, 1984, p. 305)

These conventions are unusual. They seem to be unique to van Kampen's presentation and do not appear in any other presentation of thermodynamics and it seems strange that entropy's dependence on the mole number should be determined by these propositions while no comparable constraints seem to be present for determining entropy's dependence on any other variable. What, in van Kampen's mind, makes the mole number so special in this case? We get a hint of what he is thinking at the very beginning of his article:

> "The dependence of the entropy on the number of molecules can never be found from studying closed systems." (van Kampen, 1984, p. 303)

And later, when stating his version of Gibbs' paradox, he writes:

> "[The Gibbs paradox] is a consequence of the way in which the entropy depends on the number of molecules." van Kampen (1984, p. 304)

No other commentator on the thermodynamic Gibbs paradox (except for those who explicitly endorse van Kampen's view, of course) states the root of the paradox to be in entropy's dependence on N.

This is difficult to square with the formulation of thermodynamics stated in Chapter 2 in which it is simply assumed that the entropy is a homogeneous function of the extensive variables, N being one of them. Is van Kampen honestly toiling when geometric thermodynamicists are stealing? It is difficult to say. On the one hand, it is clear that van Kampen is approaching this problem from a very experimental and operational perspective, and is therefore at pains to ask: how is it that entropy's dependence on N is determined? And he is right: from an experimental point of view, one may only determine the dependence of the dependent variable when you vary the independent variable. But, on the other hand, this is not just true of N. Why does the same argument not apply to temperature or pressure?

On the basis of these conventions he argues for the following expressions for the entropy, along with their associated interpretations.

- **The entropy of one gas:**

  $$S(p,T,N) = \frac{5}{2}NR\ln T - NR\ln p + cN \tag{VK1}$$

  where this is just Equation (4.2.3) with the constant of integration chosen to be such that $C = cN$ for some constant c.

- **The entropy of two gases, A and B:**

  $$S(p_A,p_B,T,N_A,N_B) = \frac{5}{2}(N_A+N_B)R\ln T - N_AR\ln p_A - N_BR\ln p_B + c_AN_A + c_BN_B \tag{VK2}$$

  where $p_A = pN_A/(N_A+N_B)$ and $p_B = pN_B/(N_A+N_B)$ are the 'partial pressures' of the gases A and B.

  This equation is derived by calculating $S(p_A,T_A,N_A) + S(p_B,T_B,N_B)$ where each of the summands is given by (VK1) and setting $T = T_A = T_B$.

- **The entropy of two gases, A and B:**

  $$S(p_A,p_B,T,N_A,N_B) = \frac{5}{2}(N_A+N_B)R\ln T - (N_A+N_B)R\ln(p_A+p_B) + c(N_A+N_B). \tag{VK3}$$

  This equation is derived from Equation (VK1) by setting $N = N_A + N_B$ and $p = p_A + p_B$.

We notice immediately that while Equations (VK2) and (VK3) are intended to apply when we have two gases A and B, it is not clear, from what has been said so far, when we apply one and when we apply the other. Van Kampen argues that Equation (VK3) applies when the gases are 'the same' or 'identical' and Equation

(VK2) applies when the gases are ‘different’. We shall summarise his argument for this shortly, but first we must understand van Kampen’s version of the Gibbs paradox.

The paradox stems from the fact that Convention 1 together with Equations (VK2) and (VK3) give rise to a contradiction. Convention 1 may be reasonably interpreted as a necessary condition for two systems to be ‘identical’: they are assigned the same C. If we apply this to Equation (VK2) by setting[7] $c = c_A = c_B$ then it would seem to give, according to van Kampen’s analysis, the entropy for two ‘identical’ gases A and B. But this entropy is not the same as that given by Equation (VK3), which is also supposed to give the expression for the entropy of two ‘identical’ gases. The difference between Equation (VK3) with $c = c_A = c_B$ and Equation (VK2) is

$$-N_A R \ln N_A - N_B R \ln N_B + (N_A + N_B) R \ln(N_A + N_B). \tag{DIF}$$

The preceding argument is what van Kampen takes to establish the Gibbs paradox:

> “The fact that [Equation (VK2)] does not reduce to [Equation (VK3)] when $A = B$ constitutes the Gibbs paradox. The difference is given in [Equation (DIF)].” (van Kampen, 1984, p. 306)

We are in a situation where we have a valid argument leading to a contradiction. The task, then, is to isolate the false premise.

Some readers may note that, despite van Kampen (1984) being the “go-to” reference for modern accounts of Gibbs’ paradox in thermodynamics, his account

---

[7] Despite the fact that van Kampen does not say this explicitly, it is necessary to make the step to get his result.

differs from some in a noteworthy respect. In particular, van Kampen identifies the paradox as lurking in Equation (DIF), which is the difference between the entropies of Figures (4.1c) and (4.1d), whereas more modern accounts locate the paradox in the difference between the entropies of Figures (4.1a) and (4.1c). In other words, van Kampen calculated the entropy difference between a mixture and a pure ideal gas, while modern accounts calculate the entropy difference between the unmixed state of a mixture and the mixed state.

One can formulate this observation as two slightly different puzzles. Van Kampen's puzzle is to try to ground the entropies of ideal gas mixtures and pure ideal gases in some assumption that may be interpreted as difference and identity. The other puzzle is to explain why we do not get an entropy increase when we remove a partition between two identical samples of gas. Both puzzles can be solved. In fact, once the solution to the first is found, the solution to the second drops out instantly. I will therefore focus my efforts on formulating a precise definition of identity relevant to Gibbs' paradox. But first, we must understand van Kampen's solution to the paradox.

4.2.2 Van Kampen's Solution

The logical essence of van Kampen's solution is to block the application of Equation (VK2) to identical gases. This would remove the contradiction. Van Kampen's strategy is to explain why Equation (VK2) applies only when the gases are different and (VK3) only when the gases are identical.

First, regarding identical gases he writes:

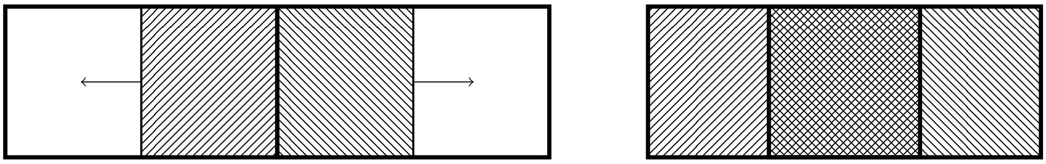
> "Consider a set of systems, consisting of identical vessels containing

> *the same gas* but in different amounts. One is free to represent all possible states obtained in this way in a space with coordinates p, T, N. There is no way, however, to compare entropy values belonging to different N, unless one introduces a new kind of process by which N can be varied in a reversible way. This is done by the familiar device of taking two identical vessels with the same p, T, N and opening up a channel between them. Together with convention 1 this leads to the conclusion that $S(p, T, N)$ must be proportional to N." (van Kampen, 1984, p. 305, original emphasis)

He thus arrives at Equation (VK1). As explained above, this allows the derivation of Equation (VK2) for non-identical gases by substituting $N = N_A + N_B$ and $p = p_A + p_B$. He does not carry out this calculation explicitly, but this step is necessary for deriving his Equation (DIF). He then goes on to consider different types of gases, writing:

> "If two vessels contain *different gases* A, B the entropy of the combined systems is defined by convention 2. When a channel is opened up an irreversible mixing occurs, so that this device cannot be utilised to define the entropy $S(p, T, N_A, N_B)$ of the resulting mixture. In its stead one now uses the familiar reasoning based on semi-permeable walls and obtains for the ideal gas [Equation (VK2)]" (van Kampen, 1984, p. 306, original emphasis)

The 'familiar reasoning' based on semi-permeable membranes is not familiar to all, so let me explain what van Kampen might have meant.



(a) Allow each gas to undergo a free expansion.

(b) Then 'overlap' the containers by applying Gibbs' theorem.

Figure 4.3: Representing mixing as two free expansions and an application of Gibbs' theorem.

Suppose before the partition is lifted we have gas A on one side and gas B on the other. And suppose for simplicity that the number of moles of each gas is the same, and that each occupy half the total volume of the container, so that the pressure of each gas is the same. We now allow each gas to undergo a free expansion into two separate volumes (see Figure 4.3a), and then 'overlap' the volumes of gas (Figure 4.3b). This chain of reasoning is a typical way of representing mixing and calculating the associated entropy increase[8].

The first step is simply two free expansions (also known as Joule expansions). The second step is an application of what is sometimes called *Gibbs' theorem.* It states that the entropy of a mixture of ideal gases at temperature T and volume V is the sum of the entropies the individual components would have if they each occupied separate containers of volume V at temperature T. Typical 'proofs' of this theorem[9] make use of the concept of *semi-permeable membranes.* These are membranes which are impermeable to one component of a mixture but permeable to all others. Installing two such membranes in between the containers of gas, one

---

[8] The most explicit example of this reasoning may be found in Adkins (1983, p. 217) but may also be found in Planck (1903, pp. 219–220).

[9] A classic text demonstrating Gibbs' theorem with semipermeable membranes is Planck (1903, pp. 219–220) and other expositions may be found in Buchdahl (1966, pp. 170–171) and Mandl (1988, pp. 129–130). Elsewhere (Wills, 2021), I argue that the correct proof of Gibbs' theorem need not rely on any concept like semipermeable membranes and is simply a mathematical consequence of the thermodynamic definition of an ideal gas mixture.

permeable to gas A but impermeable to gas B and another impermeable to gas A but permeable to gas B, allows us to overlap the gases in the very same volume and form a mixture without doing any work and while remaining at a constant temperature.

Claiming to have derived equation (VK2) via this reasoning, van Kampen then resolves the paradox by arguing that

> "two different processes had to be chosen for extending the definition of entropy. They are mutually exclusive: the first one cannot be used for two different gases and the second one does not apply to a single one." (van Kampen, 1984, p. 306)

The two processes he is alluding to here are the removal of the partition between the two samples, and the two free expansions plus the 'overlapping' with semipermeable membranes. These, it is argued, are mutually exclusive; the removal of the partition will not work for calculating the entropy of a mixture since this change of state would happen irreversibly under this process. On the other hand, the process of reversible mixing with semi-permeable membranes will not work with identical gases, since there is no differentiating feature which would allow semi-permeable membranes to exist in principle. This means that the step applying Equation (VK2) to identical gases is blocked, since the process used to establish that equation cannot be used on identical gases. This purportedly removes the contradiction at the heart of GP1.

Van Kampen's resolution has been echoed and endorsed in virtually every subsequent treatment of Gibbs' paradox in thermodynamics. His comments have been interpreted as indicating that distinguishability, in the sense that there in principle exist semipermeable membranes which allow the application of Gibbs' theorem, is what is important in resolving Gibbs' paradox. Gases may be different in any sense you like, i.e., there may be a number of properties that the gases do not share, but if there is no process which can reversibly mix or separate the gases in principle, then Equation (VK3) is to be used. Otherwise, if the gases are distinguishable in that sense, then Equation (VK2) gives the correct value for the entropy of the mixture.

There are variations on this theme of semipermeable membranes. Van Kampen actually ends up articulating a rather strong variation himself, to give the issue a pragmatic dimension: an experimenter dealing with two samples of gas may even know that they are distinguishable, but chooses to ignore this fact and treat them as if they were indistinguishable. In this case, he argues, the experimenter will be led to no contradiction in using (VK3). His solution can be summed up as the view that

> [t]he expression for entropy [. . .] depends on whether or not [the experimenter] is able an willing to make a distinction. (van Kampen, 1984)

The lesson van Kampen wants us to learn from Gibbs' paradox, is that identity (understood as the sharing of intrinsic properties) is not relevant for calculating the entropy of mixing, but rather that the more operational concept of distinguishability is what is important. This operational theme aligns with Bridgman (1961, pp. 168–169) who uses Gibbs' paradox as a case study of how his idea of the 'universe of operations' to operationally define physical concepts such as distinguishability can be used to solve problems. Another variation is Dieks (2011) who adds another level of justification, ultimately grounding the difference between (VK2) and (VK3) in differences in the physical or chemical properties between the gases, and arguing that this is sufficient to establish the existence in principle of semipermeable membranes for the pair of gases.

In summary, it is widely agreed that distinguishability via semipermeable membranes is what explains the choice between equations (VK2) and (VK3) and therefore solves Gibbs' paradox in thermodynamics.

4.2.3 Two difficulties

This account is beset by two problems. Firstly, the criterion of identity proposed by van Kampen in terms of semipermeable membranes is not fully general: it only applies to gases, when we may want to express the identity of other types of thermodynamic system. I grant that this was not van Kampen's aim: his aim was to formulate a criterion for fluids so it is no failure on his part that his criterion is not fully general. However, if we are modelling gases as thermodynamic systems, then I argue that the criterion of identity should be 'moved one level up' so we have one criterion of identity for thermodynamic systems which can be applied to fluids, rather than many criteria for all the different kinds of physical systems we may wish to model as thermodynamic. My argument in this regard is that we should seek to generalise van Kampen's methodology to thermodynamic systems in general rather than one particular case.

Secondly, this criterion of identity falls foul of the physics good practice with which we began this Chapter: this identity relation between thermodynamic models cannot be represented in the thermodynamic formalism. The definition of identity in terms of the existence of semipermeable membranes (or any reversible separation procedure) is extra-theoretic and non-thermodynamic. While the membranes are certainly useful heuristic devices for imagining reversible mixing and separation, they are not a part of the thermodynamic formalism. The criterion applies only insofar as we can supplement thermodynamics with a theory of semipermeable membranes. Although it is plausible that we do have this in the context of fluids[10], it relies on non-thermodynamic talk of the gases' constituent atoms and their atomic properties. Indeed, it is precisely because the semipermeable membrane mechanism is explicated in terms of the well-known underlying microphysics of the gases that it is so heuristically useful. But this same fact makes its usefulness totally unclear when we do not yet understand a system's microphysics, and yet are still able to model it in thermodynamics.

I should reiterate a point I made in the introduction to this Chapter: I am not making the claim that extra-theoretic reasoning is inherently bad or useless and that we should purge it from our physics. Far from it: such reasoning is immensely heuristically useful. However, extra-theoretic definitions do not help us when we are trying to prove a thermodynamic proposition such as the one stating the entropy change on mixing gases. I maintain that to be able to state and prove such a proposition, the assumptions in the argument must be thermodynamic (or, for whichever theory it is, the assumptions must appeal only to notions and formalism of the theory) and this is simply not the case when we define identity in terms of semipermeable membranes.

Thermodynamics is an extremely powerful and general theory for when we do

---

[10] For example, Planck (1903, p. 218), Zemansky and Dittman (1997, p. 387) and Partington (1913, p. 272) mention the separation of hydrogen from air using a palladium or platinum membrane in support of the the theoretical use of semipermeable membranes in thermodynamics.

not have access to the underlying microphysics of physical systems. As it happens, in the case of gases, we do know about the physics of the micro-constituents and this may be why we are naturally attracted to expressing identity in terms of the properties of those constituents. But, in general, we should not assume that we have access to the microphysics and so we should not let definitions of concepts such as identity to be tied to it. We should strive to represent the identity relation between thermodynamic models in purely thermodynamic terms.

Therefore, in spite of its popularity, van Kampen's discussion of the paradox is left wanting. Happily, there is another solution, which I turn to in the following sections.

4.3 Definition of mixtures as composite systems

Let us now use the geometric formulation of thermodynamics to make precise the notion of 'composite system', and then subsequently characterise mixtures as composite systems. While many commentators on thermal physics draw attention to the distinction between composite system and subsystem, this relationship is rarely made precise. For example, Callen (1960, p. 23) writes: "Given two or more simple systems, they may be considered as constituting a single *composite system*." He does not spell out exactly how they may be considered as such, and does not say precisely what the mathematical relationship is between the composite system and the subsystem. In this Section, I will adopt the definition of composite system from Chapter 2 which will allow us to state the precise sense in which the components of a mixture are different from each other.

To recap: suppose we have k thermodynamic models $\mathcal{M}^{(1)}, \ldots, \mathcal{M}^{(k)}$, whose manifolds are each given coordinates $(U^{(i)}, V^{(i)}, N^{(i)})$ for each $i = 1, ..., k$. I propose to define the composite of these individual thermodynamic models as follows.

Composite System. A *composite system* is the $(n{+}1)$-tuple $(\mathcal{M}^{(1)}, \ldots, \mathcal{M}^{(n)}, \mathcal{M})$ where $\mathcal{M}^{(1)}, \ldots, \mathcal{M}^{(n)}$ are n thermodynamic models whose manifolds $M^{(i)}$ may be given global coordinates $X_1^{(i)}, \ldots, X_{\dim(M^{(i)})}^{(i)}$ for $i = 1, \ldots, n$ and $\mathcal{M}$ is a thermodynamic model whose manifold M can be given global coordinates $X_0, X_1, \ldots, X_m$ each of which is a function of the coordinates $X_1^{(i)}, \ldots, X_{\dim(M^{(i)})}^{(i)}$ for all i. For each $i = 1, \ldots, n$, we say that $\mathcal{M}^{(i)}$ is a *subsystem* of the composite system.

Specifying that the coordinates of the composite system must be functions of the coordinates of the subsystems incorporates an important assumption about the relation between the composite system and its subsystems, namely that the composite system supervenes on its subsystems: there may be no change in the composite system without some change in the subsystems. I do not claim that this is the only definition of composite system possible—there may well be others—but I do claim that it relies on a plausible thesis about the relation between composite systems and their subsystems. We can use this definition can help us understand mixtures as a thermodynamic model that contains (at least) two subsystems, which represent the components of the mixture. But first we need to understand how to represent an ideal gas mixture in thermodynamics.

In Chapter 3, I concluded that an ideal gas mixture of k components is best represented by the thermodynamic model $\mathcal{M}_G$ whose manifold M_G has global coordinates $S, U, V, N_1, \ldots, N_k$ and the fundamental relation is

$$S = f_G(U, V, N_i, \ldots, N_k) = \frac{1}{T}U + \frac{p}{T}V - \sum_i^k \frac{\mu_i}{T} N_i \tag{4.3.1}$$

Variable	Definition
S	$\sum_i S^{(i)}$
U	$\sum_i U^{(i)}$
V	$V := V^{(1)} = \cdots = V^{(k)}$
N_i	$N^{(i)}$ for $i = 1, ..., k$

Table 4.1: The variables of the composite system (the ideal gas mixture) defined in terms of the variables of the subsystems (the pure ideal gases).

where

$$\frac{1}{T} = \frac{3}{2}R\frac{(\sum_i^k N_i)}{U}; \quad \frac{p}{T} = R\frac{(\sum_i^k N_i)}{V}; \quad \frac{\mu_i}{T} = R\ln\left[\frac{N_i\left(\sum_i^k N_i\right)^{3/2}}{VU^{3/2}}\right] + K_i. \quad (4.3.2)$$

This representation of an ideal gas mixture captures an important intuition we have about mixtures: they consist in a number of components in the very same volume V. We are now in a position to understand the ideal gas mixtures as a composite system in the sense described above. Let the subsystems $\mathcal{M}^{(i)}$ for $i = 1, \ldots, k$ be ideal gas models. And let the variables of the composite system (which we call the mixture of the two subsystems) be defined as in Table 4.1. A detailed physical justification for these definitions is given in Appendix A.1.2.

With these definitions, the system $\mathcal{M}_G$ satisfies the composite system definition given above, and thus we can understand more precisely the sense in which an ideal gas mixture is a composite system. We can now progress to the stage of understanding the sense in which the components of a mixture are ‘different’. I will begin this stage with a brief discussion of the identity relation and its representation in physical theory.

4.4 Representing the Identity Relation

Our careful formulation of what it means to be a thermodynamic model and what it means to be a mixture can now be put to work. I will begin by discussing how identity of physical systems is typically represented, explaining what I take to be the traditional and intuitive view of identity often appealed to in thermodynamics and in discussions of Gibbs' paradox. I show that this identity concept cannot cope with the case of ideal gas mixtures and I go on to propose a criterion of identity which can, by giving new and rigorous derivations of van Kampen's equations (VK2) and (VK3). Not only is this discussion of quite general interest, given that similar discussions concerning the identity or equivalence of systems already occur in the context of other physical theories, but it is also of crucial importance in the analysis of Gibbs' paradox.

4.4.1 The Isomorphism Criterion of Identity

On a traditional and intuitive view, the criterion of identity for physical systems is based on a more general criterion found in the philosophy of identity: two objects x and y are identical[11] when for every property P, object x has P if and only if y has P. This may be seen as a definition schema, allowing us to express a criterion of identity in more specific fields; we just need to say what the 'objects' are and what the domain of their properties is. In our case, the objects are physical systems and the domain of properties is the set of physical properties. Thus, applying the schema, the criterion of identity for physical systems is: two

---

[11] Sometimes this definition is associated with other words such as 'indistinguishability' or 'indiscernibility'. 'Identity' is a word sometimes reserved for the logical notion 'one and the same'. However, there is no consensus on the meaning of these terms and there is no scope in this chapter to discuss the intricate relationships between these similar notions.

physical systems x and y are identical when for every physical property P, object x has P if and only if y has P. This criterion, in turn, is directly imported into thermodynamics. Thermodynamic models represent physical systems, so the argument goes, so thermodynamic models are identical if and only if they share all physical properties. This move is often implicitly made in discussions of Gibbs' paradox in thermodynamics. The view is tempting because it seems to capture the intuitions we have when we say that a box of helium is 'of the same kind' as another box of helium but not of the same kind as a box of hydrogen.

However, this criterion of identity for thermodynamic models is problematic. To see why, consider the case of the difference between the boxes hydrogen and helium. It is true to say that that these substances do not have the same proton number: hydrogen's is 1, helium's is 2. And therefore it is true to say that, as physical systems, they are not identical. But, what we are interested in here is not the identity of the physical systems themselves but the identity of the physical systems *represented as thermodynamic models*. This distinction is extremely important, since it is possible for the physical systems themselves to be different while their representations as thermodynamic models are identical. This will be the case when the properties which distinguish them as physical systems are not capable of representation in thermodynamics. In our example with hydrogen and helium, thermodynamics does not have the capacity to represent the proton number of the substances and so it is possible for identical thermodynamic models to represent non-identical physical systems.

On a basic level, the distinction I am invoking here is simply that between the world and our representations of the world. I think that these are being conflated in discussions of identity in thermodynamics. Our concern should be, not the identity relation between the physical systems themselves, but the identity relation between the thermodynamic representations of those physical systems.

One may object that every difference between the physical systems themselves would manifest as a difference in their thermodynamic representations, in such a way that it would be impossible for two physical systems to be different while their thermodynamic representations are identical. In the context of our example, one may argue that the difference in proton number would ‘filter up’ and manifest as a difference in the thermodynamic models of those systems.

I agree that this is possible, but it is not necessarily so. While thermodynamics does not have the capacity to represent properties like proton number, it is possible to represent the effects of these microscopic properties ‘filtering up’ to the the thermodynamic level. An example of this is when we represent gases as van der Waals models. In our formalism, the representation of a gas such as hydrogen as a van der Waals model is $\mathcal{M}_{\mathrm{VdW}}$ whose manifold is given global coordinates S, U, V, N and whose fundamental relation[12] is

$$S/N = R\ln(V/N - b) + cR\ln(u + aN/V) + k, \tag{4.4.1}$$

where a, b, c are constants that will in general depend on the physical and chemical properties of the gas, such as the proton number, and k is the constant of integration. The constant a is meant to capture the change in energy due to intermolecular forces, the constant b is meant to capture the volume that each molecule of the gas occupies and c is the specific heat of the substance. Because of the presence of these constants in the fundamental relation, hydrogen and helium will be

---

[12] Note that this is not the more frequently discussed equation of state of the van der Waals gas: $p = RT/(V-b) - a/V^2$. It may, however, be easily derived from this fundamental relation.

different *qua* van der Waals gas because they will have different values for these constants.

This is just one way of representing the thermodynamic effect of microscopic properties when thermodynamics does not have the capacity to represent the properties themselves. However, it is not necessary for us to do this. We need not try to represent every physical property of a substance when we construct its thermodynamic model: we could just as well have represented the box of hydrogen as an ideal gas. This thermodynamic model would be $\mathcal{M}_{\mathrm{H}}$ whose manifold is four dimensional with global coordinates S, U, V, N and whose fundamental relation is that of the ideal gas. It is clear that this thermodynamic model does not have the capacity to represent any property such as proton number or any of its thermodynamic consequences; the model only states the relation between the hydrogen's entropy, volume, mole number and energy. Because of this, if a box of helium and a box of hydrogen have the same S, U, V, N, then they are identical *qua* ideal gas and different *qua* van der Waals gas.

We can be much more precise here. When we say that $\mathcal{M}_{\mathrm{He}}$ and $\mathcal{M}_{\mathrm{H}}$ have 'the same' manifolds and fundamental relations, we mean that these thermodynamic models are isomorphic: their manifolds are the same dimension and so we can define a diffeomorphism $\psi : M_{\mathrm{He}} \to M_{\mathrm{H}}$, and their fundamental relations are 'the same' in the sense that $f_{\mathrm{H}} = \psi_* f_{\mathrm{He}}$. On the other hand, the boxes of hydrogen and helium *qua* van der Waals gas are not isomorphic structures since their fundamental relations are 'different' in virtue of the differing values of the constants a, b, c. This leads us to state a criterion of identity in terms of the isomorphism of the formal structures:

The Isomorphism Criterion of Identity. thermodynamic models $\mathcal{M}$ and $\mathcal{M}'$ are identical iff:

- There exists a diffeomorphism $\psi : M \to M'$ and
- $f = \psi_* f'$

where M, M', f and f' are the manifolds and fundamental relations of the thermodynamic models $\mathcal{M}$ and $\mathcal{M}'$ respectively.

Isomorphism is typically held in high regard as a natural standard of equivalence[13] and, as an equivalence relation, is a good candidate for representing identity between two physical systems represented by formal structures in some theory. In our case, the isomorphism of the ideal gas structures captures an important sense in which the hydrogen and helium physical systems are 'the same' at room temperature and pressure: their physical behaviour is the same in this regime and is represented extremely well by the ideal gas model.

But, this leads us to a curious conclusion in the context of thermodynamic mixtures and Gibbs' paradox. If we regard isomorphism of the thermodynamic models as representing the identity relation between the physical systems, then it follows that all ideal gases are identical. Therefore, it would be impossible to have an ideal gas mixture, whose very definition relies on the ideal gas components of the mixture being different in some sense. The isomorphism criterion of identity fails to capture the common practice of modelling two ideal gases as non-identical (such as in Gibbs' paradox). If this is a modelling practice we wish to keep, then we need to find a way of modelling non-identical ideal gases.

---

[13] See, for example, Weatherall (2018, p. 331) for an application of this criterion in the context of general relativity.

The strategy I adopt is inspired by considering mixtures; two different gases combined will form a mixture, while two identical gases combined is just more of the same gas. The strategy is to convert this innocuous fact into the definition of identity: two gases are identical when they combine to form 'more of the same stuff' and are different when they don't. In this sense, the identity question has been changed: rather than picking up any two thermodynamic models and asking whether they are identical (as we could with the isomorphism criterion), we should instead ask whether two thermodynamic models are identical if they are each part of a third system which is the composite of the two. How the composite system is defined yields the identity relation between the subsystems: if the composite is a mixture, that is to represent the subsystems as non-identical, if the composite is more of the same stuff, that just is to represent the subsystems as identical. This section will execute this strategy in detail. I begin by providing rigorous derivations of van Kampen's equations (VK2) and (VK3). This will expose crucial assumptions in the derivations which I will go on to generalise and interpret as difference and identity.

4.4.2 Deriving equations (VK2) and (VK3)

Van Kampen's Equations (VK2) and (VK3) purport to be the entropy of a mixture and an ideal gas respectively. In this Section, I show how they may be derived from the fundamental relations of composite systems where the subsystems with variables $S^{(i)}, U^{(i)}, V^{(i)}, N^{(i)}$ are pure ideal gases.

We begin by deriving Equation (VK2). This follows from the fundamental relation of the ideal gas mixture, given in Section 4.3. Again, the fundamental relation for the ideal gas mixture is

$$S = f_G(U, V, N_1, ..., N_k) = \frac{1}{T}U + \frac{p}{T}V - \sum_i \frac{\mu_i}{T}N_i, \tag{4.4.2}$$

where

$$\frac{1}{T} = \sum_i \frac{3}{2}R\frac{N_i}{U}; \tag{4.4.3}$$

$$\frac{p}{T} = \sum_i R\frac{N_i}{V}; \tag{4.4.4}$$

$$\frac{\mu_i}{T} = R\ln\left[\frac{N_i\left(\sum_j N_j\right)^{3/2}}{VU^{3/2}}\right] + K_i \text{ for } i = 1, ..., k. \tag{4.4.5}$$

By eliminating U in favour of T, defining the 'partial pressures' $p_i := N_i/V = pN_i/(\sum_i N_i)$, and absorbing all constants into the K_i, we recover van Kampen's Equation (VK2) for k components in the following form:

$$S = \frac{5}{2}\left(\sum_{i=1}^{k} N_i\right) R\ln T - \sum_{i=1}^{k}\left(N_i R\ln p_i\right) + \left(\sum_{i=1}^{k} N_i K_i\right) \tag{VK2$'$}$$

where the summation indices run over the number of components in the mixture.

We now turn to deriving Equation (VK3). Since it purports to be the entropy of a 'mixture' of identical gases A and B, this 'mixture' will be a pure ideal gas of the same kind as A and B. So, the coordinates of M^{VK3} are going to be those of the ideal gas: S, U, V, N. The definitions of the variables for $\mathcal{M}^{\mathrm{VK3}}$ are summarised in Table 4.2 and their physical justification is given in Appendix A.1.3.

The fundamental relation for the ideal gas is

Variable	Definition
S	$\sum_i S^{(i)}$
U	$\sum_i U^{(i)}$
V	$\sum_i V^{(i)}$
N	$\sum_i N^{(i)}$

Table 4.2: Definitions of the variables for $\mathcal{M}^{\mathrm{VK3}}$ for deriving Equation (VK3).

$$S = f^{\mathrm{VK3}}(U, V, N) = \frac{1}{T}U + \frac{p}{T}V - \frac{\mu}{T}N, \tag{4.4.6}$$

with the equations of state

$$\frac{1}{T} = \frac{3}{2}R\frac{N}{U}; \tag{4.4.7}$$

$$\frac{p}{T} = R\frac{N}{V}; \tag{4.4.8}$$

$$\frac{\mu}{T} = R\ln\left[\frac{N^{5/2}}{VU^{3/2}}\right] + K. \tag{4.4.9}$$

If we eliminate U in favour of T and V in favour of p, and define $p^{(i)} := N^{(i)}RT/V$ while also writing $N = \sum_i N^{(i)}$ and $p = \sum_i p^{(i)}$ we recover van Kampen's (VK3) in the following form:

$$S = \frac{5}{2}R\left(\sum_i^k N^{(i)}\right)\ln T - \left(\sum_i^k N^{(i)}\right)R\ln\left[\sum_i^k p^{(i)}\right] + \left(\sum_i^k N^{(i)}\right)K \tag{VK3$'$}$$

Thus, we have shown how to recover van Kampen's equations (VK2) and (VK3) from the fundamental equations of an ideal gas mixture and the pure ideal gas respectively, and by treating van Kampen's gases A and B as subsystems of a composite.

At this point, we may compare our Equations (VK2$'$) and (VK3$'$) with van Kampen's (VK2) and (VK3) and notice an interesting and important feature of the interpretation of the indices. Van Kampen uses A and B, which are supposed to label the gases, to index the variables in both equations. But what is referred to with 'the gases'? What is the relation of A and B to the composite system in each case?

What we learn from these derivations is that use of the indices A and B is ambiguous. It is at least clear that they are meant to be subsystems of the composite in some sense, but in what sense, precisely? Here, I have shown that the gases can be subsystems of the composite in two different ways, as illustrated in Tables 4.1 and 4.2. In the next section, we will investigate to what extent these senses can be interpreted as difference and identity. But at the very least we can note here that van Kampen's use of the same indices in each equation covers up the fact that the indices take on quite different interpretations. Let us start with comparing Equation (VK2$'$) and van Kampen's (VK2). In our (VK2$'$), the subscript indices i run over the components in the mixture. However, while each N_i is a proper thermodynamic variable in the sense that each is a coordinate of M, the so-called 'partial pressures' p_i are not thermodynamic variables in the sense that they are not partial derivatives of f_G with respect to any configuration variable.[14] In (VK3$'$) the indices run over subsystems. In this case, neither the $N^{(i)}$, the mole number of subsystem (i), nor the $p^{(i)}$, the pressure of subsystem (i) is a proper thermodynamic variable since the coordinates of M are S, U, V, N. Therefore, while van

---

[14] Callen (1960, p. 337) emphasised this point. He writes "The partial pressures are purely mathematical constructs with no direct physical meaning." This is a little cryptic, but he may be taken to mean that these pressures are not derivatives of the fundamental relation with respect to any configuration variable.

Kampen wrote his entropies as functions of p_A, p_B, T, N_A, N_B for both (VK2) and (VK3) respectively, they really should be written as functions of p, T, N_A, N_B and p, T, N respectively.

In this Section, we have given rigorous derivations of each of van Kampen's equations. This has clarified the status and interpretation of the indices A and B used by van Kampen in his expression of the equations. We have learned that it is not as simple as "letting A and B represent two gases" which may or may not be identical; we have to specify the precise nature of the subsystem-composite system relationship before interpreting the indices. The purpose of examining detailed derivations of these equations is precisely so that we may examine whether any of the assumptions behind the equations may be interpreted as saying that the systems $\mathcal{M}^{(i)}$ are 'different' or 'identical' in some sense. We now turn to this issue.

4.4.3 The Homogeneity Criterion of Identity

Here I propose a criterion of identity for thermodynamic models which allows isomorphic ideal gas models to be modelled as non-identical, paving the way for making sense of ideal gas mixtures, while also having much more general and foundational implications. Van Kampen's Equation (VK2) purports to be the entropy of a mixture of 'different' substances while (VK3) purports to be the entropy when the substances are 'the same'. With our derivations of Equations (VK2$'$) and (VK3$'$), we have uncovered the assumptions behind these equations and in this section I will argue that these assumptions can be interpreted as 'difference' and 'identity' respectively.

In order to state the precise criterion of identity for thermodynamic models, we must introduce some formal machinery. Given a chart (U, φ) on M where $U \subseteq M$ and $\varphi : U \to \mathbb{R}^k$, let us adopt the notation $(U, \lambda\varphi)$ to denote the chart that 'scales' each of the coordinate components of $\varphi(U) \in \mathbb{R}^k$ by λ, i.e. $\lambda\varphi$ is defined by $x_i \circ \lambda\varphi(p) := \lambda x_i \circ \varphi(p)$ for each coordinate component $x_i : \mathbb{R}^k \to \mathbb{R}$ for $i = 1, .., k$ and for all $p \in U$.

The Homogeneity Criterion of Identity. Thermodynamic subsystems $\mathcal{M}^{(i)}$ for $i = 1, \ldots, n$ in a composite system $(\mathcal{M}^{(1)}, \ldots, \mathcal{M}^{(n)}, \mathcal{M})$ for $i = 1, \ldots, n$ are identical iff:

- There exist diffeomorphisms $\psi^{(i)} : M^{(i)} \to M$ for all i;
- If, for each (i) and for any chart (U, φ) on M, $({\psi^{(i)}}^{-1}(U), \varphi^{(i)})$ is a chart on $M^{(i)}$ such that $\varphi^{(i)} = \lambda\varphi \circ \psi^{(i)}$ (that is, $\varphi^{(i)}$ is the scaled chart pushed forward to $M^{(i)}$) then:

$$\left(f^{(i)} \circ {\varphi^{(i)}}^{-1}\right)\left(X_1^{(i)}\left({\psi^{(i)}}^{-1}(p)\right), \ldots, X_k^{(i)}\left({\psi^{(i)}}^{-1}(p)\right)\right) = \lambda\left(f \circ \varphi^{-1}\right)(X_1(p), \ldots, X_k(p)) \tag{ID}$$

  for all $p \in U$, and where $X_j := x_j \circ \varphi$ and $X_j^{(i)} := x_j \circ \varphi^{(i)}$ for $j = 1, \ldots, k$ are the coordinate maps on M and $M^{(i)}$ respectively.

This criterion of identity for thermodynamic models is fully general and specifies a criterion for each element of the ordered pair. The first part of the criterion requires that the relata of the identity relation be subsystems of some composite. The second requires (at least) that the manifolds be of the same dimension. The third says that for $\mathcal{M}^{(i)}$ for $i = 1, \ldots, n$ to be identical as subsystems of the composite $\mathcal{M}$ means that if $X_j^{(i)} = \lambda X_j$ for $j = 1, \ldots, k$ for some positive real λ then we have that $f^{(i)}(X_1^{(i)}, ..., X_k^{(i)}) = \lambda f(X_1, ..., X_k)$. This is placing a constraint on the behaviour of the fundamental relations when the coordinates $X_j^{(i)}$ and X_j for $j = 1, \ldots, k$ are related by a scaling transformation, namely the fundamental relations also must be related by the same scaling transformation. To see how this criterion relies on the homogeneity assumption, take the fundamental relation of the composite system $f(X_1, \ldots, X_k)$ and scale each of the coordinates thus: $f(\lambda X_1, \ldots, \lambda X_k)$. The fundamental relation of subsystem (i) is $f^{(i)}(X_1^{(i)}, \ldots, X_k^{(i)})$. The condition for identity is that their fundamental relations are the same when $X_j^{(i)} = \lambda X_j$ for $j = 1, \ldots, k$, so we have $f^{(i)}(X_1^{(i)}, \ldots, X_k^{(i)}) = f(\lambda X_1, \ldots, \lambda X_k)$. By homogeneity, this is $f^{(i)}(X_1^{(i)}, \ldots, X_k^{(i)}) = \lambda f(X_1, \ldots, X_k)$.

The basic idea that this criterion tries to capture is the following: if we want to represent the proposition that k physical systems are identical in thermodynamic terms, this is to say that when they are composed, the resulting composite system is a scaled version of each of the subsystems. Thus, the representation of the identity relation is based on the relation between a composite system and its subsystem and the assumption of homogeneity.

It may be objected at this point that I have said nothing useful; I may be interpreted as saying that systems are different if they form a mixture. But is whether or not they combine to form a mixture exactly what we are trying to establish? Isn't the problem we are trying to solve when we have a mixture and when we do not?

In short, no. I am engaged in the project of trying to *represent* the identity relation between systems in thermodynamics. To elaborate: I make no claim about the 'correct' way to distinguish between systems. We may distinguish them by their atomic properties, or by our abilities to reversibly separate them or by arbitrarily calling one ‘A’ and the other ‘B’. My question then is, having distinguished them (or not) by one’s chosen means, how does one represent this identity or difference in thermodynamics? My answer is given above. This is not to be read as a ‘test’ of whether systems are identical or not but as a way of representing thermodynamically the result of that test. From this point of view, I have no objection to van Kampen’s operational way of distinguishing ideal gases by their method of separation. What I am objecting to, on the other hand, is his thermodynamic representation of the claim that the gases are different or identical. In this way, my account is perfectly compatible with this aspect of van Kampen’s argument and, indeed, complements it.

We will discuss the philosophical implications of this criterion after I show that it does the job: to precisely drive a wedge between Equations (VK2) and (VK3) and hence solve GP1. This directly addresses the contradiction that is at the heart of van Kampen’s puzzle. Using this criterion of identity, I show that the subsystems of the system $\mathcal{M}_G$ are different to their composite, and the subsystems of the system $\mathcal{M}^{\mathrm{VK3}}$ are identical to their composite.

We begin by showing that any subsystem $\mathcal{M}^{(i)}$ of $\mathcal{M}^{\mathrm{VK3}}$ is identical to $\mathcal{M}^{\mathrm{VK3}}$. We note first that the manifolds $M^{(i)}$ and M^{VK3} are of the same dimension and hence (given also standard smoothness conditions) there exists a diffeomorphism between them. In defining the coordinates of the composite system, we found the relation between them to be $\lambda_i X = X^{(i)}$ for $X = U, V, N$ and where $\lambda_i := N^{(i)}/\sum_i N^{(i)}$. This is just to say that the coordinates of the systems are scaled versions of each other. Therefore, since the fundamental relations are first-order homogeneous, we have $f^{(i)}(U^{(i)}, V^{(i)}, N^{(i)}) = f(\lambda_i U, \lambda_i V, \lambda_i N) = \lambda_i f(U, V, N)$. What we have shown is that each of the subsystems $\mathcal{M}^{(i)}$ is identical to the composite system $\mathcal{M}^{\text{VK3}}$. In the case of VK2, the identity condition for the manifold fails: the dimension of M_G is not the same as the dimension $M^{(i)}$ so there can be no diffeomorphism between them. Thus, the composite system is not identical to any of the subsystems. We may look at the difference between VK2 and VK3 a different way. For VK2, the mole numbers $N^{(i)}$ of the subsystems were defined to be the mole numbers N_i of the individual components of the composite system. It is essentially this which means that the substances i whose quantities are denoted by the variables N_i are all different: they are independently variable components and hence form a mixture when combined. In contrast, the mole numbers $N^{(i)}$ of the subsystems of $\mathcal{M}^{\text{VK3}}$ were defined to be part of the total mole number $N = \sum_i N^{(i)}$. It is in this sense that the composite system is just "more of the same kind of stuff" as the subsystems $\mathcal{M}^{(i)}$.

As we saw, the isomorphism criterion captures an important sense in which ideal gases can be identical but it fails to make sense of ideal gas mixtures. The scaling criterion was developed in response to this, providing a way in which ideal gases can be different as components of a mixture. It is natural to question the relationship between these two criteria. Do we need both? If so, when do we use which criterion?

Apart from the case of ideal gas mixtures described in this paper, there are more general and fundamental reasons for the introduction of the scaling criterion: it allows us to represent isomorphic thermodynamic models as non-identical. Why might we want to do this in general? Recall that thermodynamics is a theory which deliberately ignores some (possibly known) physical properties of a physical system. This is another way of expressing the typical characterisation of thermodynamics as focusing only on the macroscopic physics of a system while abstracting away from the microphysics. It follows that when thermodynamics represents a physical system as a thermodynamic model, it does not represent some properties which the physical system has, or is known to have. The ideal gas is the perfect example of this; ideal gases with different chemical composition and which differ in a great number of physical properties are all represented by the very same thermodynamic model. But what if we want to 'thermodynamically acknowledge' isomorphic thermodynamic models as being different in virtue of properties which thermodynamics cannot represent? For example, is there a way of representing hydrogen and helium as different in virtue of their physical properties, while still representing them as isomorphic ideal gas models? The scaling criterion allows us to do this by considering the hydrogen and helium as subsystems of a composite. We may represent the composition of isomorphic ideal gas models as an ideal gas mixture which means that each subsystem fails to scale to the composite. In this way, the 'non-scaling' of each isomorphic subsystem to the composite allows us to acknowledge all the physical properties which the subsystems do not share but which thermodynamics does not have the capacity to represent.

Ultimately, when to use the isomorphism criterion and when to use the scaling criterion will depend on the context and one's modelling goals. But we may briefly sum up the different use cases of these two criteria as follows. The isomorphism criterion involves at least two individual thermodynamic models and it represents their identity in terms of their individual formal structures. The scaling criterion involves at least three systems: at least two subsystems, and their composite. The identity of the subsystems in this case depends on how the composite system is defined. For example, if we choose to represent the composition of two isomorphic ideal gas models as an ideal gas mixture, then we are representing the isomorphic ideal gas models as different since they each do not scale to the composite. If, on the other hand, we represent the composition as an ideal gas, then the isomorphic subsystems do scale to the composite and hence they are identical (in both senses). While the scaling criterion involves at least three thermodynamic models and the full structure of the subsystem-composite system relation, it is able to represent isomorphic thermodynamic models as different. This is desirable when we want to acknowledge two isomorphic thermodynamic models as different in virtue of properties which thermodynamics does not have the capacity to represent.

4.5 Conclusion

This Chapter has introduced a new way of representing the identity relation in thermodynamics which goes beyond the definition in terms of isomorphism. This new notion provided a new perspective on Gibbs' paradox, which arose because we lacked a thermodynamic representation of the identity relation which could account for the intuition that the components of a mixture are different.

Chapter 5

Classical Particle Indistinguishability

5.1 Introduction

Considerations of the identity of objects have long been part of philosophical discussion in the natural sciences and logic. These considerations became particularly pertinent throughout the twentieth century with the development of quantum physics, widely recognized as having interesting and far-reaching implications concerning the identity, individuality, indiscernibility, indistinguishability[1] of the elementary components of our ontology. This discussion continues in both the

---

[1] This menagerie of terms in the literature is apt to cause confusion, especially as there is no clear consensus on what exactly each of these terms mean. Throughout the rest of the paper, I will be concerned with giving a precise definition of a concept which I will label 'indistinguishability', and which, I think, matches with how most other commentators use the word. But I believe it to be a distinct debate whether particles are 'identical', 'individual', or 'indiscernible'. The literature I address in this paper therefore does not, for example, overlap with the debate over the status of Leibniz's PII.

physics and philosophy literature[2].

Such considerations in classical physics have, by comparison, been largely ignored. This trend has rightly been resisted in the philosophical and foundational literature, but discussion of the idea has remained small in comparison to the (perhaps more glamorous?) discussion in quantum theory. It is strange that this is the case, since considerations of particle indistinguishability in physics actually first made their debut in Gibbs' work on classical statistical mechanics (Darrigol, 1991, p. 239). Gibbs used the notion of particle indistinguishability to argue for the introduction of a factor $N!$ in the probability distribution over the phase space for a system of N 'entirely similar' particles. It is here where the problem of giving a rigorous justification for the $N!$ factor originates.

The question of whether, and if so in what precise sense, classical particles are indistinguishable has consequences in physics and its philosophy: arguments which aim to justify the inclusion of the $N!$ factor generally assume that it is needed to derive an extensive entropy function. The basic worry is that a non-extensive entropy function yields a non-zero entropy of mixing of indistinguishable samples of gas, contradicting the prediction of thermodynamics that it is zero. This contradiction is one version of Gibbs' paradox[3], and seeking a justification for the $N!$ factor is often taken to be synonymous with seeking a solution to the paradox. The argument I will present follows the trend in the philosophical literature by taking seriously the idea of indistinguishable classical particles, but does not take as its basic motivation the derivation of an extensive entropy function. In fact, I

---

[2] See French (2000) for an introduction to this discussion and a comprehensive list of references.

[3] See van Kampen (1984), Denbigh and Redhead (1989), Jaynes (1992), Dieks (2011), Dieks (2013), Saunders (2013), van Lith (2018), and Swendsen (2018) for a sample of the debate from the perspective of seeking an extensive entropy function.

think it is a virtue of my analysis that it shows that a rigorous derivation of the $N!$ factor can be obtained on a very general level, independent of any particular probability measure and independent of what we use the probability measure to compute. This shows that the $N!$ problem is not a problem specific to the entropy, but rather a deep puzzle in the foundations of statistical mechanics.

However, Gibbs' treatment of indistinguishable particles is beset by ambiguous and imprecisely defined terms (what exactly does it mean for classical particles to be 'indistinguishable'?). This lack of clarity has resulted in a surprisingly diverse collection of justifications for the $N!$ factor. Some prominent textbook authors like Huang (1963), ter Haar (1966), and Blundell and Blundell (2010) appeal to quantum mechanics, arguing that the insertion of the $N!$ in classical physics was necessary but not justifiable until quantum mechanics came along with its associated 'fundamental indistinguishability of identical particles'. Indeed, much of the modern discussion and teaching of the physics of indistinguishable particles occurs in the context of quantum mechanics. In contrast, philosophers and those working in the foundations of physics have supplied arguments which defend the coherence of the idea of the indistinguishable classical particle. For example, Saunders (2013) and Saunders (2018) supplies a justification of the $N!$ factor on the basis of a notion of classical particle indistinguishability as defined by the intrinsic properties of classical particles like mass and charge. On the other hand, Dieks (2013) argues that the 'distinguishability' of classical particles means that the $N!$ should not be included after all.

In this paper, I present an analysis of classical particle indistinguishability as 'observational indistinguishability' in a certain mathematically precise sense. I will argue that this leads to a number of interesting and welcome consequences in the foundations of statistical mechanics:

1. The identification and resolution of shortcomings in the ongoing debate concerning the solution to the $N!$ problem.

2. A reinterpretation of the quotienting procedure typically used to justify the $N!$ term and a rigorous derivation of the $N!$ factor which does not appeal to the metaphysics of particles and which rather draws only on facts about observables.

3. A reinterpretation and reconstruction of Gibbs' own argument as a special case of my analysis in which particles are observationally indistinguishable with respect to the Hamiltonian. I call this 'dynamical indistinguishability'.

In Section 5.2, I present the necessary formalism and concepts of classical statistical mechanics in order to state a precise formulation of the $N!$ problem. In Section 5.3, I critically evaluate two prominent approaches to the $N!$ factor debate. In Section 5.4, I demonstrate that considerations of observability are crucial to understanding the sense in which states may be said to be 'equivalent' for the purpose of gathering statistics, and I offer a new perspective on the nature of reduction via quotienting in classical mechanics. In Section 5.5, I offer a reconstruction of Gibbs' argument for which I develop a precise definition of classical particle indistinguishability in terms of the symmetries of the dynamics. In Section 5.6, I use this definition to finally derive the $N!$ factor.

5.2 Gibbs' Statistical Mechanics

In classical statistical mechanics[4], the main object of study is the *Hamiltonian system.* This is a triple $\langle M, \omega, h\rangle$ where M, a smooth manifold representing the possible states of the system, together with ω, the symplectic form supplying M with geometric structure, form the *phase space.* Given a *Hamiltonian*, $h : M \to \mathbb{R}$, the symplectic form ω determines a unique dynamical trajectory through each point in phase space. The dynamical trajectories are the integral curves of the vector field H generated by h via the symplectic form thus:

$$w_{ab}H^a = d_b h \tag{5.2.1}$$

where $d_b h$ is the exterior derivative of h. We interpret the different integral curves with various initial conditions as representing the possible dynamical trajectories the system may follow from the various initial states represented by those points. Dynamical trajectories may also be viewed as maps $\varphi_t : M \to M$ which take points in M and output points in M, t units along the integral curve from the initial point. The collection of these maps $\{\varphi_t\}_{t\in\mathbb{R}}$ forms a one-parameter group of diffeomorphisms and is called the *phase flow.*

When the points in the phase space for a system of n degrees of freedom are given local coordinates $(q_1, ..., q_n, p_1, ..., p_n)$, where $q_1, ..., q_n$ are the generalized configuration variables and the $p_1, ..., p_n$ are the generalized momenta, Equation

---

[4] I choose to present the Gibbs rather than the Boltzmann version of statistical mechanics. This is partly because the problem I am concerned with first originated with Gibbs. Despite this, many authors in the modern philosophical literature (like Saunders (2013) and Dieks (2013) which I discuss below) choose to state the problem in the Boltzmann framework. Happily though, this does not affect the solution of the problem since, ultimately, all the action occurs in the probability measure over the phase space which is a feature common to both Gibbs and Boltzmann versions.

(5.2.1) takes on the more familiar form of Hamilton's equations:

$$\dot{q}_i = \frac{\partial h}{\partial p_i},\ \dot{p}_i = -\frac{\partial h}{\partial q_i} \text{ for } 1 \leq i \leq n. \tag{5.2.2}$$

Now we introduce the statistical part. In introducing his statistical mechanical framework, Gibbs invites us to

> imagine a great number of systems of the same nature, but differing in the configurations and velocities which they have at a given instant. (Gibbs, 1902, p. iii)

This collection of systems is known as the *ensemble*: an enormous collection of copies of the system under study which contains all of the possible ways the actual system could be at an instant. It is important to recognize that the ensemble is not (for example) the collection of particles in the actual box of gas itself; it consists of possible, not actual, copies of the entire system. And it is the ensemble, not the actual system, that is the object of study in Gibbs' statistical mechanics:

> And here we may set the problem, not to follow a particular system through its succession of configurations, but to determine how the whole number of systems will be distributed among the various conceivable configurations and velocities at any required time, when the distribution has been given for some one time. The fundamental equation for this inquiry is that which gives the rate of change of the number of systems which fall within any infinitesimal limits of configuration and velocity. (Gibbs, 1902, p. iii)

He is not concerned with the traditional mechanical problem of how a particular state of the system changes in time but rather with the problem of how the distribution of a number of systems over the possible states changes with time.

We are invited to imagine that the ensemble of systems is distributed in a particular way among the possible mechanical states in phase space. This distribution D on phase space tells us how many systems N in an ensemble are associated with a given region R of the phase space, through the following definition:

$$N(R) := \int_R D(p,q)\, dp_1 ... dq_n. \tag{5.2.3}$$

Gibbs calls $D(p,q)$ the *density-in-phase* and $dp_1 ... dq_n$ an *element of extension-in-phase*[5].

The total number of systems in the ensemble $N := N(M)$ is given by Equation (5.2.3) for $R = M$. Gibbs normalizes D by defining $P(p,q) := D(p,q)/N$, which Gibbs calls the *probability coefficient*[6]. He then defines

$$\mu(R) := \int_R P(p,q)\, dp_1 ... dq_n. \tag{5.2.4}$$

We then define

$$\mathcal{P}(R) := \frac{\mu(R)}{\mu(M)} \tag{5.2.5}$$

to be the probability of region R. This defines the probability measure of a region $R \subseteq M$. Its intended interpretation is as the probability that a system picked out of the ensemble is found in region R of the phase space. In Gibbs' words, it

---

[5] In modern language, this is known as the 'volume form' or 'Liouville form' associated with the symplectic manifold (M, ω), given by $\omega^n = dq_1 \wedge dp_1 \wedge \cdots dq_n \wedge dp_n$.

[6] In modern language, this is a 'probability density function'.

expresses

> [the] ratio of the number of systems falling within those limits to the whole number of systems. This is the same as the probability that an unspecified system of the ensemble [...] will lie within the given limits. (Gibbs, 1902, p. 17)

This quotation suggests that the talk of ensembles of great numbers of systems is just a concrete way of interpreting the probability measure in a frequentist manner. However, in using Gibbs' formalism, we are not committed to the existence of this ensemble of systems[7]. The important message is that the central object of study in Gibbsian statistical mechanics is the probability measure over the phase space. Physics proceeds by computing expectation values for certain observables which are then compared with experiment.

Much of Gibbs' 1902 book studied the probability coefficients of arbitrary Hamiltonian systems of n degrees of freedom interacting with different environments. For example, the first probability coefficient he studied was the canonical distribution. He interpreted this to model systems which could exchange energy with their constant-temperature environment. He then introduced the microcanonical distribution to study the properties of systems with a fixed energy. If a physicist wants to study the properties of their particular system when it is in contact with a heat bath (for example), all they have to do is take the canonical distribution, which Gibbs has already worked out and apply it to their particular system, effectively endowing the canonical coordinates with certain interpretations.

---

[7] Mathematically, all we have is a probability distribution over M. How we are to interpret this probability is a matter of debate. See Myrvold (2017) for a discussion of the main issues.

For example, for a single point particle with three degrees of freedom, the generalized configuration variables represent the particle's x, y and z position coordinates and the generalized momenta represent the particle's linear momentum along the x, y and z axes. It is at this stage where the problem we are concerned with arises.

While Gibbs spent most of his book studying arbitrary systems of n degrees of freedom, in the last chapter of his book he did apply his methods to a particular system: a collection of N point particles[8]. In applying his methods to this system, he expressed the following problem:

> If two phases differ only in that certain entirely similar particles have changed places with one another, are they to be regarded as identical or different phases? (Gibbs, 1902, p. 187)

This quote requires some interpretation. The $6N$-dimensional phase space of a system of N point particles has phase points (also known as phases) s of the form of a $6N$ tuple:

$$s = (r_{1x}, r_{1y}, r_{1z}, p_{1x}, p_{1y}, p_{1x}, ..., r_{Nx}, r_{Ny}, r_{Nz}, p_{Nx}, p_{Ny}, p_{Nz}) \tag{5.2.6}$$

Or, more briefly

$$s = ((\mathbf{r}, \mathbf{p})_1, ..., (\mathbf{r}, \mathbf{p})_N) \tag{5.2.7}$$

---

[8] This last chapter is known for the grand canonical distribution, used in modelling systems of particles in which the particle number is allowed to vary (in chemical reactions, for example). Indeed, it is in the mathematical expression of the grand canonical distribution where the problem I discuss first appeared. However, the discussion and solution to the problem will not depend on any particular probability distribution; the problem would appear if we applied the canonical, micro-canonical or another distribution to the N particle system.

where $\mathbf{r} = (r_x, r_y, r_z)$ and $\mathbf{p} = (p_x, p_y, p_z)$. We are to think of this tuple as consisting of N 'slots', known as *factor positions*, each of which represents one of the N particles, and into which go three position coordinates and three momentum coordinates specifying the single particle state. This phase point represents the state of the N particle system, made up of N single particle states. Now consider the family of transformations $\{\pi_{ij}\}$ for $1 \leq i, j \leq N$ whose elements swap the blocks of six canonical coordinates, which is to say they swap the single-particle states, among the factor positions i and j. For example, the transformation $\pi_{1N} : M \rightarrow M$ is defined so that

$$\pi_{1N}(s) := ((\mathbf{r}, \mathbf{p})_N, ..., (\mathbf{r}, \mathbf{p})_1). \tag{5.2.8}$$

Let us refer to these transformations as *permutations*. The permutations π_{ij} effectively change which particle has which position and momentum. It is easily checked that these permutations are canonical transformations, a fact which will be useful to us later.

The permutation π_{ij} is sometimes described as a 'permutation of particle labels' (French and Rickles, 2003) or as a 'permutation of particles among the states' (Dieks, 1990; Saunders, 2013). These images are tempting, but they can be misleading as to what the permutations $\{\pi_{ij}\}$ really do. Firstly, we cannot think of particles as being permuted among states since this invites the image of phase points remaining fixed, like boxes, and particles being permuted between them. In fact, the formalism indicates that we should think of the particles like boxes (they are represented by the factor positions) as remaining fixed while the states (blocks of six coordinates) are permuted among them. Nor can the maps be interpreted passively as a permutation of particle labels since, strictly speaking, there are no labels for the map to act on; all that the N tuple consists in is factor positions and single particle states. It is possible to interpret the labels as names referring to the factor positions (the label 'A' would just be shorthand for 'the first factor position in the tuple' and so on), and if we must talk of labels then this is how they should be interpreted, but this still does not capture what the maps do since, again, the factor positions remain fixed. As defined, these maps act on N tuples and return N tuples by permuting single particle states among the factor positions. The π_{ij} do not permute particle labels over particles, or particles over states: they permute states over particles[9].

Now Gibbs poses the following question: what if the particles represented by factor positions i and j are 'entirely similar' or 'indistinguishable'? The transformation π_{ij} certainly takes s to a different phase point $\pi_{ij}(s) \in M$, $s \neq \pi_{ij}(s)$, but are the states represented by $\pi_{ij}(s)$ and s, in some sense, equivalent? It is tempting to say in this case that if the particles are 'indistinguishable', it does not 'matter' which particle has which single particle state and so the states represented by the phase point s and its permuted counterpart $\pi_{ij}(s)$ are in some sense equivalent. This is a common interpretation of Gibbs' reasoning here, for he writes:

> If the particles are regarded as indistinguishable, it seems in accordance with the spirit of the statistical method to regard the phases as identical. (Gibbs, 1902, p. 187)

---

[9] The distinction between permuting particles over states and permuting states over particles has made made in the context of quantum theory: French and Krause (2006, p. 136) note that we may think about permutation both ways in the context of ordinary statistics but that the distinction becomes important in the context of parastatistics. In this paper, however, we restrict ourselves strictly to classical Hamiltonian mechanics for a contained discussion.

If all N particles were 'indistinguishable' from each other, then it seems that the phase space is over-representing the physical states; each physical state is represented by $N!$ points in the phase space since N single particle states over N factor positions admit of $N!$ distinct permutations of single particle states over particles. This has a knock-on effect on the probability measure: if the physical states are over-represented in M, the standard probability measure over M developed in the earlier part of Gibbs' book for arbitrary systems would be over-counting the states. Gibbs corrects for this[10] by dividing the probability measure by $N!$. More precisely, if μ is the measure derived from the general reasoning of earlier chapters (see Equation (5.2.4)), the measure for the N particle system would be $\mu/N!$ to correct for the over-counting of the states by μ.

The terms Gibbs uses are open for interpretation: what is meant by 'entirely similar' or 'indistinguishable'? What does it mean to say that the phases are 'identical'? Is this to be read as logical identity? Or some other notion of equivalence? These ambiguities have led to a number of views regarding the $N!$ factor. In the next Section, I critically evaluate two prominent views, taking certain parts of their arguments as inspiration for my own derivation of the factor and for my answers to these questions.

---

[10] In the original text, Gibbs considers the more general case in which the gas consists of numbers of particles of different kinds so his factor is $N_1!N_2!...N_h!$ where $1, ..., h$ indexes the different kinds of particles in the gas such that particles of kind 1 are 'indistinguishable' from each other but 'distinguishable' from any other kind. We work with the simpler case of N particles of the same kind for simplicity and without loss of generality.

5.3 $N!$ Justifications

In this Section, I turn to discussing and critically evaluating the arguments of Dieks and Saunders, two prominent representatives of different views regarding the $N!$ factor. This will help set the scene for the alternative view I will defend in this paper.

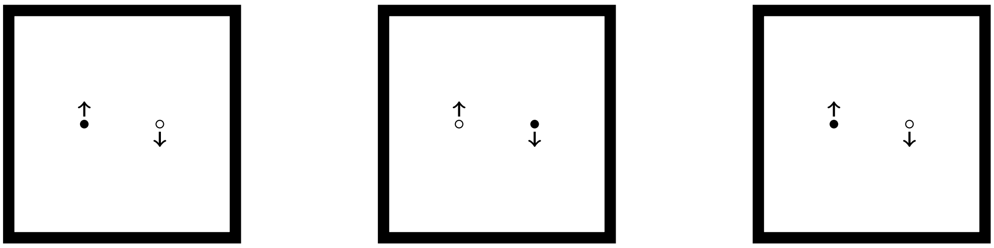
5.3.1 Dieks' Classical Particle Distinguishability

Dennis Dieks gives an argument against the inclusion of the $N!$ factor, based on the idea that classical particles cannot be indistinguishable at all because they are always distinguishable by their trajectories in configuration space. The thrust of his argument is expressed here:

> [C]lassical particles can be named and distinguished by their different histories. A process in which two classical particles of the same kind are interchanged can therefore certainly produce a different microstate. Indeed, imagine a situation in which there is one particle at position $[r_1]$ and one particle at position $[r_2]$, and in which at a later instant there is again one particle at $[r_1]$ and one at $[r_2]$; suppose that their respective momenta are the same as before. What has happened in the meantime? There are two possibilities: either the particle that was first at $[r_1]$ is later again at $[r_1]$ and the particle that was first at $[r_2]$ is later again at $[r_2]$, or the particles have exchanged their positions. The latter case would clearly be different from the former one: it corresponds to a different physical process. Although it is true that the two final situations cannot be distinguished on the basis of their instantaneous

> properties, their different histories show that the particle at $[r_1]$ in one final situation is not the same as the particle at $[r_1]$ in the other final situation. (Dieks, 2013, p. 27)

The situation Dieks has in mind here is depicted in Figure 5.1. Let the initial state (a) be represented by the phase point $s = (r_1, p_1, r_2, p_2)$. And let state (b) be the final state of the two particles if they swap, taking the system to the state $\pi_{12}(s) = (r_2, p_2, r_1, p_1)$, where $\pi_{12} : M \to M$ is the permutation, acting on the single particle states in the factor positions 1 and 2. This transformation represents the process in which the particle (represented by the first factor position, coloured black in the diagram) which initially had (r_1, p_1) now has (r_2, p_2) and the particle (represented by the second factor position, coloured white in the diagram) which initially had (r_2, p_2) now has (r_1, p_1). The particles have swapped single particle states. State (c) shows the final state of the two particles if they do not swap, the system returning to the state s. The puzzle here is that if the particles are 'of the same kind', 'entirely similar', or 'indistinguishable' then the



(a) Initial state (b) Final state (swap) (c) Final state (no swap)

Figure 5.1: (a) is the initial state of a system consisting of two particles in a container. The arrows indicate their velocities. States (b) and (c) are the final states of particles after they follow trajectories which lead them to swapping and not swapping, respectively. The black dot depicts the particle represented by the first factor position and the white dot depicts the particle represented by the second factor position.

states $\pi_{12}(s)$ and s are in a sense equivalent, since in each there is one particle at position r_1 with momentum p_1 and another at position r_2 with momentum p_2; physically, it might not seem to matter which particle has which position and momentum. Dieks argues that although the states are not distinguishable via their 'instantaneous properties', the particles in Figure 5.1b followed different trajectories in configuration space to those in 5.1c. So, when we take the histories into account, it does seem to 'matter' which particle is which. In this sense, the particles are clearly distinguishable via their trajectories and therefore the states $\pi_{12}(s)$ and s are, in another sense, not equivalent. Which is the correct sense of equivalence to take? Dieks argues that the histories are relevant to considerations of the equivalence of the states and so he concludes that the states are distinct.

In order to evaluate the argument, let me put it in a more formal setting. Note first of all that the system in which the particles do not swap must have a different dynamics than the system in which the particles do swap. More formally, representing the dynamics as the phase flow after some fixed time $\varphi_t : M \to M$, then it must be the case that:

$$\varphi_{\neg\text{swap}} : s \mapsto s \tag{5.3.1}$$

$$\varphi_{\text{swap}} : s \mapsto \pi_{12}(s) \tag{5.3.2}$$

In other words, if the system is to evolve under the dynamics to some phase point other than s, then the phase flow must be different. Each phase flow is generated by the corresponding Hamiltonian h^{swap} and $h^{\neg\text{swap}}$. But the Hamiltonian is a function of the phase—indeed, it is the energy—and thus is exactly the sort of object that may fairly be called an 'instantaneous property' of the system. Therefore, the two final states in Figures 5.1b and 5.1c can in fact be distinguished on the basis of their instantaneous properties, since the Hamiltonian in each case is different. It is exactly because the states are different in virtue of their different Hamiltonians that the states have different trajectories in phase space.

Therefore, what Dieks has really told us is that the phase point s in the phase space of dynamical system $\langle M, \omega, h^{\neg\text{swap}}\rangle$ is a different state from $\pi_{12}(s)$ in the phase space of dynamical system $\langle M, \omega, h^{\text{swap}}\rangle$. However, this is not the question we are really trying to answer. What we want to know is whether s in the phase space of some dynamical system is in some sense equivalent to $\pi_{12}(s)$ in the phase space of *the very same dynamical system.* What are we to say about the sense in which the states represented by s and $\pi_{12}(s)$ are equivalent?

5.3.2 Saunders' Classical Particle Indistinguishability

Saunders (2013) and Saunders (2018) supplies an answer to this question as part of an argument justifying the inclusion of the $N!$ factor in the context of Boltzmann statistical mechanics. If the particles are identical[11], understood to mean that the particles share all state-independent properties (like charge, mass and spin), then the states $\pi_{ij}(s)$ and s (for $1 \leq i, j \leq N$) represent one and the same physical state (the sense of equivalence used here being identity). This representational redundancy of the phase space then needs to be removed by passing to the reduced phase space, $M' = M/S_N$: the quotient of the phase space under the action of

---

[11] This term is not meant in the strict logical sense of 'one and the same', but it is widely used, so I adopt it. In the philosophical literature, particle identity has been treated with significant sophistication and precision. See, for example, Caulton and Butterfield (2011), Caulton and Butterfield (2012), and French (1989).

S_N, the permutation group[12]. Points in the reduced space may be understood as equivalence classes $[s]$ each containing $N!$ elements all related to the $s \in M$ by the permutations $\pi \in S_N$. Quotienting under this group effectively identifies points in the unreduced phase space related by these permutations. More informally, the unreduced phase space points are read as saying, ‘the first particle has this position and momentum, the second particle has that position and momentum, and so on’ while the points in the reduced phase space are read as saying ‘some particle has this position and momentum and another particle has that position and momentum, and so on’ While it is possible to make statements about which particle has which position and momentum in the unreduced space, it is not possible to say which is which in the reduced space.

This reduction has the desired effect on the probability measure. Let there be a measure μ on the unreduced space M such that $\mu(M) = 1$ (it is normalized) and let μ' be the measure on the reduced space M' obtained through the quotienting procedure. We may embed M' to M by mapping the points $[s] \in M'$ to some point $[s] \ni s_i \in M, 1 \leq i \leq N!$. This maps M' onto some region $R \subset M$. Carrying μ' over to M then allows allows us to compare μ' and μ on the same space. Since $\mu(R) = 1/N!$, and R can be mapped to the entire reduced space, it follows that $\mu'(M') = \mu(R) = 1/N!$ resulting in the desired $N!$ factor.

There are two key steps in this argument that I will focus on. These are:

1. Focusing on the key question of whether the points in M related by some $\pi \in S_N$ represent ‘one and the same’ physical state. The criterion for this was taken to be: the points s and $\pi_{ij}(s)$ represent one and the same physical

---

[12] S_N contains $N!$ elements which are generated by the transformations π_{ij}.

state just in case the particles represented by factor positions i and j are identical.

2. If the states are equivalent in this sense, then the points s and $\pi_{ij}(s)$ are to be identified by quotienting M under the action of the permutation group.

I take issue with the first, and therefore am not convinced by the second. My critique is the subject of the next Section. In its place I will propose a new and mathematically precise perspective on the equivalence of states, which in turn gives rise to a novel interpretation of the quotienting procedure.

5.4 Observables and Statistics

If we assume that the job of the state space in a typical dynamical theory is to represent completely and without redundancy all the possible ways a system can be at an instant, then the question in step one of whether two states are 'one and the same' physical state is pertinent. The situation in statistical mechanics, however, is slightly different. While we are ostensibly studying a dynamical system, it is not the primary object of study. This is because statistical mechanics is a statistical theory and, as such, it is not immediately concerned with the evolution of the state of the system, but rather with studying a *probability space*. This is a triple $(\Omega, \mathcal{F}, \mu)$ where Ω is the sample space whose elements are called outcomes, $\mathcal{F}$ is a σ-algebra on Ω whose elements are called events and $\mu : \mathcal{F} \rightarrow [0, 1]$ is a countably additive probability measure. This shift in focus is important: while the job of the state space M is to completely represent all the possible states of the dynamical system, the job of the subsets of Ω is to be the events to which probabilities are attached. In the previous section, it was tacitly assumed that Ω was M, but this will not, in general, be the case, as I will show shortly. It is important to realize that M and Ω are conceptually different entities with different roles vis-à-vis what they aim to represent. The importance of noting this difference and its relevance to statistical mechanics is the topic of this section.

In a typical statistical analysis using a probability space, the job of Ω is to represent the set of possible outcomes or results of an experiment or observation while $\mathcal{F}$ represents the set of possible events which are assigned probabilities (c.f. Billingsley, 1995, p. 17). I would like to draw attention to the seemingly innocuous role of observation in defining the sample space. The appearance of observation in the standard interpretation of a probability space is, I think, very telling and important. We can only collect statistical data and determine the relative frequencies of various elements of $\mathcal{F}$ when the outcomes in the sample space are observationally distinguishable from one another. Therefore, considerations of what counts as observable must be accounted for when specifying the sample space. To give a really simple example: we have two coins to be flipped at the same time, each of which can either land heads (H) or tails (T) and with one coloured red (R) and the other green (G). If the experimenter tossing the coins and recording relative frequency data is colourblind, it will only be possible for them to collect statistical information with the sample space $\{(H,H),(H,T),(T,T)\}$, even though the state space is in fact $\{(GH,RH),(GT,RH),(GH,RT),(GT,RT)\}$. That is, the sample space is constrained by the observables the experimenter has available. With respect to the observables available to the colourblind experimenter, the states (GT,RH) and (GH,RT) are observationally indistinguishable.

Given that the primary object of study in statistical mechanics is the probability space, we should be in the business of constructing the sample space. And given that observational considerations are inextricably bound up in the standard characterisation of a sample space, we must pay close attention to the observable in constructing the sample space for the N particle system. These considerations of what is observable will help to pin down the relevant notion of state equivalence to use in constructing our sample space: we should not be asking whether two points represent one and the same state but rather whether they are observationally equivalent. This is an important difference, since states may be observationally equivalent while not being one and the same state[13]. We therefore need a precise characterisation of observational equivalence of states in classical mechanics.

Observable quantities in classical mechanics are represented by classical observables. These are smooth, real valued functions on the manifold $f : M \to \mathbb{R}$, assigning a real number to each state. This allows us to define precisely what it means for two states to be observationally equivalent.

Observational Equivalence. States s and s' are *observationally equivalent* with respect to a set of observables $\{f_i\}$ if and only if $f_i(s) = f_i(s')$ for all f_i.

This definition provides a schema for constructing the sample space. Simply identify your favourite set of smooth functions and build your sample space containing all and only states which are observationally distinct (inequivalent) with respect to that set.

We may consider a close cousin of Saunders' view as an instance of this schema

---

[13] Note that, in constructing the sample space, we are not making any claims about which states are 'really' equivalent. We are in fact not modifying the phase space in any way at all, but rather building a new space, constrained by the observables, on which to subsequently define probabilities and do statistical mechanics. The argument presented here is, therefore, compatible with any empirically adequate phase space ontology.

in which we interpret his talk of particle properties in terms of observables. The next step is to interpret the claim that the particles have the same intrinsic properties in terms of classical observables. We cannot read the intrinsic properties of the particles, such as charge and mass, in the N particle system from the state since this only gives us position and momentum values. In fact, we read intrinsic properties of the particles from the classical observables. This is best illustrated with an example: if we want to say that the particles all have the same mass, we would formalize this in terms of the permutation symmetry of some classical observable, like the kinetic energy[14]. Thus, consider:

1. $K = (p_1^2 + p_2^2)/2m$;
2. $K' = p_1^2/2m + p_2^2/2M$, $m \neq M$.

The variables p_1 and p_2 pick out the momentum values in the factor positions 1 and 2 respectively. K is left invariant under the permutation π_{12} since $K(s) = K(\pi_{12}(s))$ for all s, while K' is not so invariant. We can say then that the particles represented by factor positions 1 and 2 have the same mass if and only if the kinetic energy observable is invariant under the permutation π_{12}. This is one interpretation of what it means for particles to have the same mass. Similarly, having the same value of other particle properties such as charge and spin will also be understood formally in terms of the permutation symmetry (or not) of some observable(s).

Thus, while Saunders' view says that the necessary condition for states s and $\pi_{ij}(s)$ to be identified in the quotienting procedure is that the particles share all

---

[14] This will not be the only classical observable invariant under this transformation, but it is the obvious one to consider.

state independent properties, the close cousin of Saunders' view would say that the necessary condition for the states to be identified in the quotienting procedure is that the states must be observationally equivalent with respect to all observables. This view is an instance of the schema with $\{f_i\}$ being the set of all smooth functions.

This instance of the schema has an undesirable consequence for Saunders with respect to step 2: it implies that no two distinct states in phase space are ever observationally equivalent:

Proposition. Let $s, s' \in M$, where M is a smooth connected manifold. If $f(s) = f(s')$ for all smooth functions $f : M \to \mathbb{R}$, then $s = s'$.

Proof. Suppose $f(s) = f(s')$ for all functions f, and let C be any curve on M containing s and s'. Now let f be any smooth function that increases strictly monotonically along C. Then since $f(s) = f(s')$ and s, s' are both on C, it follows that $s = s'$. $\square$

Therefore, if we require two states to be observationally equivalent with respect to all observables before identifying the states in the quotienting procedure, then quotienting will not actually do anything, since no two distinct states are observationally equivalent in this sense.

However, one could reasonably respond that to require observational equivalence with respect to all observables is an impossibly strong epistemic requirement. There is the inevitable experimental limitation that we are able to measure only a few observables. If we want a useful σ-algebra whose events we can actually observationally distinguish, then the sample space had better contain all and only the states observationally distinguishable with respect to some subset of the observables, such as those we can measure. This claim has two important implications.

The first implication affects step 2: the quotienting procedure. The goal of reduction is to remove representational redundancy in the phase space by identifying states (deemed physically identical) in the unreduced space as one state in the reduced space. That aim is sometimes viewed as metaphysical: it is to achieve a bijection between physically possible states of the system and points in the state space[15]. This, however, is not the goal of a statistical theory, such as statistical mechanics. All we require of our sample space is that the elements of Ω be distinguishable with respect to some set of empirically accessible (measurable) observables. In fact, given our observational and experimental limitations, we know that the elements of the sample space may correspond to many genuinely distinct possibilities in the state space (recall that this was the lesson of the coin example). Since we inevitably do not have access to all the classical observables, this is our situation in statistical mechanics. So, if we wish to retain the ability to interpret phase space as representing distinct physical possibilities, then reducing the phase space is not the correct procedure to carry out, since this would identify states known to be genuinely distinct.

One way around this problem is to re-interpret the reduction procedure; instead of interpreting the reduced space as a state space in one-to-one correspondence with genuine possibilities, we could interpret it as a sample space, claiming to represent all and only the observationally distinct states without claiming to represent all and only the genuine physical possibilities. If interpreted this way, the

---

[15] See Castellani (2003, p. 434), Belot (2003, p. 410) and Butterfield (2007, pp. 23–4) for examples of this view. Examples of state space reduction typically occur in the relationist-substantivalist debate in spacetime physics.

reduction procedure can be carried out in a representationally accurate way. This interpretation is non-standard since reduction is typically used to eliminate the over-representation of possibilities. But nothing in the mathematics of reduction dictates any particular interpretation of the procedure; if we think of quotienting in terms of forming equivalence classes of states in M, we are at liberty to quotient with respect to any equivalence relation which suits our purposes. Typically, this relation is taken to be '...is identical to...' or '...is physically equivalent to...' but we may just as well take the relation to be '...is observationally equivalent to...' In the rest of the Chapter, however, I will adopt a position which does not require reduction under any interpretation. This position has a number of virtues and advantages over re-interpreting reduction which I explain in Section 5.6.

The second implication is more pressing because it also affects the possible ways out of the problem posed by the first implication. Since no distinct states are observationally equivalent with respect to all observables, which subset of observables will give us a fruitful definition of observational equivalence? A fruitful definition would be one which allows us to construct a sample space which is actually useful, that is, contains all and only the outcomes which are observationally distinct with respect to observables available to an experimenter. This set is clearly dependent on the experimental context and our epistemic situation and therefore there will be no 'one true' fruitful definition of observational equivalence.

In the next Section, I will consider one particular definition of observational equivalence resulting from choosing the subset of observables to be the singleton set containing the Hamiltonian. I will show how this results in the $N!$ factor used by Gibbs and a precise definition of classical particle indistinguishability in terms of dynamical symmetries. But realize that this is only one possible way of filling in the observational equivalence schema: the division by $N!$ results from specifying a particular choice for $\{f_i\}$. Thus the '$N!$ problem' is a specific instance of the much more general 'over-counting problem'. Depending on the choice of observables, we will in general correct by a different factor.

5.5 Dynamical Equivalence

A hint that $\{f_i\} = \{h\}$ is the correct starting point for interpreting and reconstructing Gibbs is provided by Uffink. On introducing 'indistinguishable' particles, he makes the following comment (in a footnote) on what that might mean:

> Presumably, these terms mean (at least) that the Hamiltonian is invariant under their permutation, i.e. they have equal mass and interact in exactly the same way. (Uffink, 2007, p. 997)

The hint that the indistinguishability of particles is grounded in some symmetry of the energy observable (the Hamiltonian) is very promising, but it certainly requires further argument[16]. The energy observable is but one of many smooth functions on phase space: why choose this one with respect to which to judge the (in)distinguishability of the particles and the observational equivalence of the states? In the following sections I develop in detail the line of thought inspired by this hint and I show that it results in a precise definition of classical particle indistinguishability, and a rigorous derivation of the $N!$ factor.

We shall call observational equivalence with respect to the Hamiltonian[17] h in

---

[16] This hint carries through straightforwardly to the discussion in quantum theory, in the form that the permutation operator commutes with the Hamiltonian. However, I restrict this Chapter to the discussion in classical Hamiltonian mechanics.

[17] The Hamiltonian which is the third element of the dynamical system triple (M, ω, h).

which $\{f_i\} = \{h\}$, dynamical equivalence, to be defined more precisely below. But first, this view does require some motivation.

This position is motivated by the thought that a minimal, necessary condition for two states of a dynamical system to be observationally equivalent is that they have identical dynamical evolutions. It should not be possible to count states as observationally equivalent for the purposes of constructing the sample space if they have different dynamical evolutions.

Is the condition sufficient? It is clearly not sufficient for observational equivalence, since taking $\{f_i\} = \{h\}$ allows the states to differ with respect to any number of other observables, if these differences do not make any difference to their dynamical evolutions. For example, the particles may all have different colours but if the dynamics is 'colourblind', then it should not matter if this particle is blue and that one red or this one red and that one blue. They do, nevertheless, differ with respect to the 'colour' observable. Recall, however, that our aim is to find the set of observables sufficient for our statistical aim of constructing the sample space and correct probability distribution for our Hamiltonian system.

An argument that $\{h\}$ is sufficient for this purpose comes from Chapter IV in Gibbs' book in which he arrives at the canonical distribution. He begins by singling out as worthy of study the distributions P which are stationary. That is, those which satisfy

$$\dot{P} = \{P, h\} = 0 \tag{5.5.1}$$

where $\{P, h\}$ is the Poisson bracket of P and h. He notes that it is necessary and sufficient that P be a function only of the phase for this condition to be satisfied. We may further constrain P by requiring that they define a probability measure over M by satisfying Equations (5.2.4) and (5.2.5). In Gibbs' terminology, the conditions specified by Equations (5.2.4), (5.2.5) and (5.5.1) combine to form the condition for statistical equilibrium. This still leaves the functional form of P underdetermined, so Gibbs places a further constraint on P: based on the restriction of his study to conservative mechanical systems, he stipulates P to be some function of the energy:

$$P(s) = \alpha(h(s)) \text{ for some } \alpha : \mathbb{R} \to \mathbb{R} \text{ for all } s \in M. \tag{5.5.2}$$

Part of the reason for this stipulation is, no doubt, to focus discussion on particular probability coefficients; taking P to be some function of the energy is an easy and natural way to satisfy the condition for statistical equilibrium. However, it is a mathematical fact that $\{P, h\} = 0$ for a conservative system does not imply that P must be some function of the energy. Gibbs' stipulation that this be so therefore appears rather arbitrary from a philosophical point of view.

However, it is a fact that Gibbs restricted his attention to stationary probability coefficients which were some function of the energy. Insofar as we are in the business of reconstructing Gibbs' reasoning, or indeed any reasoning which restricts attention to functions of the energy, dynamical equivalence is appropriate.

I now develop a precise statement of dynamical equivalence. Consider the phase flow φ_t generated by h and the particular trajectory

$$t \mapsto \varphi_t(s) \in M \tag{5.5.3}$$

with initial condition s. Now compare the trajectory with initial condition $\pi(s)$ for some $\pi \in S_N$:

$$t \mapsto \varphi_t(\pi(s)) \tag{5.5.4}$$

We also define the permutation of the curve $t \mapsto \varphi_t(s) \in M$ as

$$t \mapsto \pi(\varphi_t(s)) \tag{5.5.5}$$

This is just like the curve $t \mapsto \varphi_t(s)$ except that every point along the curve has been acted upon by π. Intuitively, it is the very same trajectory, differing only in the factor positions of the single particle states. This motivates a definition for states s and $\pi(s)$ to be dynamically equivalent.

π-Dynamical Equivalence. States s and $\pi(s)$ are π-*dynamically equivalent* if and only if $\varphi_t(\pi(s)) = \pi(\varphi_t(s))$, for all t.

In words, this asks you to compare two trajectories and starting at some point and the permuted version of that point. If they are the same trajectory up to permutation π, then the states are π-dynamically equivalent. Abstracting away from particular trajectories, this becomes: if the transformation π leaves the hamiltonian flow invariant, then we say that the states s and $\pi(s)$ are π-dynamically equivalent.

This definition does not yet look like an instance of the observational equivalence schema in Section 5.4. However, we can prove that it is equivalent to one such instance using the following result: if a canonical transformation leaves the hamiltonian flow invariant, then this leaves the Hamiltonian invariant up to an additive constant. The geometric proof goes as follows[18].

Proposition. Given a symplectic manifold (M, ω), let $\varphi_t : M \to M$ be the Hamiltonian flow generated by the smooth function $h : M \to \mathbb{R}$. Let $\pi : M \to M$ be a symplectomorphism (canonical transformation), such that $\pi^*\omega = \omega$ where π^* is the pushforward of π. Then: $\varphi_t \circ \pi = \varphi_t$ iff $h \circ \pi = h + f_0$ where f_0 is a constant function.

Proof. $\to$. A Hamiltonian flow for h is a set of integral curves threading the vector field H. So the proposition is equivalent to the statement that $\pi^*H = H$ iff $h \circ \pi = h + f_0$. H is defined to satisfy $\omega_{ab}H^a = d_b h$. Writing the inverse of ω_{ab} as ω^{ab}, we have $H^a = \omega^{ab} d_b h$. So $\pi^*H = H$ is to say $\pi^*\omega^{ab} d_b h = \omega^{ab} d_b h$. Since π^* is a symplectomorphism, $\omega^{ab}\pi^* d_b h = \omega^{ab} d_b h$. Applying ω_{ab} to both sides, we have $\pi^* d_b h = d_b h$. All diffeomorphisms commute with the exterior derivative, so this implies that $d_b \pi^* h = d_b h$. By the definition of pullback, this implies $d_b(h \circ \pi) = d_b h$. By linearity of the exterior derivative, $d_b(h \circ \pi - h) = 0$. This implies that $h \circ \pi - h = f_0$.

$\leftarrow$. Suppose $h \circ \pi = h + f_0$. Taking the exterior derivative, we have $d_b(h \circ \pi) = d_b h$. Now, $d_b(h \circ \pi) = d_b(\pi^* h) = \pi^* d_b h$. So $\pi^* d_b h = d_b h$. Applying ω^{ab} to both sides we have $\omega^{ab}\pi^* d_b h = \omega^{ab} d_b h$. π^* is a symplectomorphism so $\pi^*\omega^{ab} d_b h = \omega^{ab} d_b h$. This is just to say that $\pi^* H = H$. $\square$

The truth of this claim is perhaps more easily seen in the coordinate setting of Hamilton's equations. Here is a sketch of how it works. Suppose the Hamiltonian h gives rise to the flow $(q_i(t), p_i(t))$. Suppose we canonically transform the coordinates so that $h \mapsto h'$ and $(q_i(t), p_i(t)) \mapsto (Q_i(t), P_i(t))$. Since the transformation

---

[18] An equivalent proposition and proof may be found in Corollary 8 in Arnol'd (1978, pp. 217–8)

is canonical, this flow results from plugging in some Hamiltonian h' into Hamilton's equations. Now suppose that this canonical transformation actually leaves the flow invariant: $q_i(t) = Q_i(t)$ and $p_i(t) = P_i(t)$, for all i, t. Then we have that $\partial h'/\partial p_i = \partial h/\partial p_i$ and $\partial h'/\partial q_i = \partial h/\partial q_i$ respectively. It follows that $h' = h + f(q_i)$ and $h' = h + g(p_i)$ respectively. This means that $h' = h + f_0$ where f_0 is a constant.

This proposition allows us to see the definition of π-dynamical equivalence as an instance of the observational equivalence schema, with $s' = \pi(s)$, $\{f_i\} = \{h\}$: states s and $\pi(s)$ are π-dynamically equivalent if and only if $h(\pi(s)) = h(s) + f_0$, for all s.

So far, we have talked about dynamical equivalence of states in a general and abstract geometrical setting. We now need to understand how these ideas on the level of the N particle state filter down to the level of the individual particles. To this end, we now apply these general geometrical ideas in the coordinate setting to better understand the functional form of certain typical Hamiltonians and their symmetries under the permutation maps. This will in turn motivate a definition of the dynamical indistinguishability of classical particles.

(a) $s = (1, -1, 2, 1)$. (b) $\pi(s) = (2, 1, 1, -1)$.

Figure 5.2: Two states related by a permutation. In each case, there is one particle with $(r, p) = (1, -1)$ and another with $(2, 1)$ and yet the states are represented by two distinct points in the phase space. The black dot depicts the particle represented by the first factor position and the white dot depicts the particle represented by the second factor position.

Consider a system of two particles moving in one dimension along the x-axis. The four dimensional phase space of this system has points of the form $(r_{1x}, p_{1x}, r_{2x}, p_{2x})$, two of which are depicted in Figure 5.2. According to the proposition just proved, the states s and $\pi(s)$ are dynamically equivalent when the Hamiltonian is left invariant (up to an additive constant[19]) by the transformation: $h(\pi(s)) = h(s) + f_0$.

This is satisfied by $h_1 = (p_1^2 + p_2^2)/2m$. Explicitly, the p_1 picks out the momentum value in the first factor position which is -1 for s. p_2 picks out the momentum value in the second factor position which is 1 for s. When the Hamiltonian acts on $\pi(s)$, p_1 picks out 1 and p_2 picks out -1 since these values now occupy different factor positions as a result of applying π to s. However, it is clear that it does not matter which value is inserted into p_1 or p_2 because of the symmetry of the Hamiltonian: $h_1(s) = ((-1)^2 + (1)^2)/2m = ((1)^2 + (-1)^2)/2m = h_1(\pi(s))$.

This condition is not satisfied, however, by $h_2 = (p_1^2 + p_2^2)/2m + GMm/r_1$, where GM/r_1 is the gravitational potential at r_1 (the position of the particle represented by the first factor position) of some fixed, stationary mass M. $h_2(s) = ((-1)^2 + (1)^2)/2m + GMm/(1)$ while $h_2(\pi(s)) = ((1)^2 + (-1)^2)/2m + GMm/(2)$ since r_1 picks out the value 1 for state s whereas it picks out the value 2 in state $\pi(s)$. It is clear that (in general) $h_2(s) \neq h_2(\pi(s))$, since the position and momentum variables pick out different numbers as a result of the permutation. The fact that h_2 is not symmetric in this sense means that the states s and $\pi(s)$ have different dynamical evolutions. s and $\pi(s)$ are dynamically equivalent with respect to h_1 but dynamically inequivalent with respect to h_2.

Let us consider a slightly more complicated Hamiltonian of a four particle system moving in one dimension: $h_3 = (p_1^2 + p_2^2 + p_3^2 + p_4^2)/2m + ke_1e_2/|r_{13}|$, where $ke^2/|r_{13}|$ is the electric potential energy term as a function of the relative

---

[19] That the Hamiltonian is only invariant up to an additive constant is not a problem, since we only ever observe changes in energy anyway.

distance between the particles represented by factor positions 1 and 3 both with charge e and k is a constant. The particles represented by factor positions 2 and 4 are neutral. In this case, the Hamiltonian satisfies $h_3(s) = h_3(\pi_{13}(s))$ and $h_3(s) = h_3(\pi_{24}(s))$ for all s but it is not the case that, in general, $h_3(s) = h_3(\pi_{23}(s))$ nor $h_3(s) = h_3(\pi_{14}(s))$. The reason for this is that particles 1 and 3 are neutral and 2 and 4 have the same electric charge, so permuting single particle states among those factor positions makes no difference to the dynamical evolution. However permuting states among factor positions 1 and 2 for example, would lead to a different dynamical evolution, since this swaps states between a particle with an electric force acting on it with a particle with no such force.

These three examples of Hamiltonians motivate a precise definition of the 'dynamical indistinguishability' of classical particles:

Dynamical Indistinguishability. Two particles of an N particle system, represented by factor positions i and j in the N tuple, are *dynamically indistinguishable* if and only if the Hamiltonian h of the system satisfies $h(s) = h(\pi_{ij}(s)) + f_0$ for all states s.

An immediate consequence of this definition is that the dynamical indistinguishability of the particles is relative to the Hamiltonian; in fixing h, we make claims about the dynamical indistinguishability of certain particles in that system. We may link dynamical indistinguishability with the more intuitive and familiar, albeit less precise, notion of identical particles. This is best illustrated with an example.

Consider a collection of particles in a gas which are all neutral except for one charged particle. This particle will, of course, not interact with any of the neutral particles. If there is no electromagnetic field present, then there will be no electromagnetic interactions at all, and the Hamiltonian will contain no electromagnetic potential terms. Although the particle is charged, this charge is not being exploited by any field; there are no forces on this particle which depend on this charge, and so the charge is irrelevant to the dynamical evolution of the system. The charged particle would be dynamically indistinguishable from the neutral particles, even though they are not identical. We may change the Hamiltonian by turning on an electromagnetic field, inducing a force on the charged particle resulting in a different trajectory than it would have otherwise taken. In virtue of this, the particle is dynamically distinguishable from the other neutral particles.

This shows that differing intrinsic properties of the particles is a necessary but not sufficient condition for dynamical distinguishability. The particle must be charged for the field to act only on it with a force, but the mere fact that it is charged is not sufficient for it to be dynamically distinguishable, since it may be treated as neutral if there is no electromagnetic field present in the system. Thus, while it is necessary that there be some difference for the particles to be dynamically distinguishable, it is not necessary that they share all their intrinsic properties for them to be dynamically indistinguishable. Particles may be dynamically indistinguishable even if they do in fact differ with respect to certain intrinsic properties, since the Hamiltonian may be 'blind' to those properties.

5.6 Solving the $N!$ Problem

In this Section, I show how we may rigorously derive the $N!$ factor on the basis of observational equivalence of states, and the observational indistinguishability of particles. I will make this argument in the particular case of dynamical equivalence, since this is the situation that Gibbs and others often assume, but it is fully generalisable. The result in this Section is a particular instance of the more general problem of deriving some over-counting correction factor on the basis of some definition of the observational equivalence of states.

Let us now consider, for simplicity, the Hamiltonian $h = \sum_{i=1}^{N} \mathbf{p}_i^2/2m$. In this case, all the $\pi_i \in S_N$ are dynamical symmetries. Thus, given a state $s \in M$, $\pi_i(s) \in M$ is dynamically equivalent to s for all $\pi_i \in S_N$. What I aim to show is that the manifold may be partitioned into $N!$ regions (for this particular hamiltonian), each of which contain all and only points which are observationally (=dynamically) distinguishable from one another and such that each region is observationally indistinguishable from any other. In this way, any member of the partition may be considered a sample space of a probability space.

To begin, note that the relation "...is dynamically equivalent to..." is an equivalence relation on M and this induces a partition on M with $N!$ member sets, each of which is an equivalence class $[p]$ of points dynamically equivalent to p. This by itself is not what we are after, since each member of the partition contains points which are dynamically equivalent to each other, whereas what we want, in order for each member of the partition to be considered a sample space, is for the points in the member of the partition to be dynamically *in*equivalent to each other. Roughly speaking, what we want is to take the "cross-section" of this partition, forming another partition in which each member contains exactly one point from each equivalence class $[p]$. In this way, we will have partitioned M into $N!$ regions U_i each of which may be considered a sample space.

We may construct this partition more precisely as follows. Define the subset[20] $M \supset U_1 := \{s \in M : \mathbf{r}_1 < \mathbf{r}_2 < \ldots < \mathbf{r}_N\}$, i.e. take all the points in M such that the position coordinates of the N factor positions are ordered in the manner specified[21]. The claim is that this U_1 is a sample space, i.e. it contains all and only points in M dynamically inequivalent from each other. To prove this, we have to show, firstly, that no two points in U_1 are dynamically equivalent to each other and secondly, that no other point in $M \backslash U_1$ is dynamically inequivalent to any point in U_1.

To prove the first, take an arbitrary point $r \in U_1$ and consider the dynamically equivalent point $\pi(r)$ for some $\pi \in S_N$. By the definition of the permutations π and the set U_1, $\pi(r) \notin U_1$ since the permutation maps swap the factor positions and hence the ordering of the position values. Therefore, no two points in U_1 are dynamically equivalent. To prove the second, suppose there is some point $s' \notin U_1$ and dynamically inequivalent to all points in U_1. Since $s' \notin U_1$, its position values are not ordered in the manner specified. But this means there is some map π such that $\pi(s') \in U_1$, i.e. s' is related to some point in U_1 by a permutation transformation. But if this is the case, then it is dynamically equivalent to some point in U_1, contrary to assumption. In sum, U_1 contains all and only points dynamically inequivalent to each other.

Having defined U_1 and shown that it has the desired properties, we generate the partition of $N!$ dynamically equivalent regions recursively:

---

[20] We have to clarify what this ordering relation means for triples: $\mathbf{r}_1 < \mathbf{r}_2$ iff $r_1^x < r_2^x$ and $r_1^y < r_2^y$ and $r_1^z < r_2^z$.

[21] We do not also need to specify an ordering on the momentum coordinates because, since the pair $(\mathbf{r}_i, \mathbf{p}_i)$ occupies one factor position, it is fixed by the ordering on the position coordinates. I also follow standard practice in assuming the impenetrability of Newtonian particles, thus ignoring the "collision" singularities where particles occupy the same position.

$$U_1 := \{s = ((\mathbf{r}_1, \mathbf{p}_1), \ldots, (\mathbf{r}_N, \mathbf{p}_N)) \in M : \mathbf{r}_1 < \mathbf{r}_2 < \ldots < \mathbf{r}_N\}$$

$$U_2 := \{s : s = \pi_1(x) \text{ for all } x \in U_1\}$$

$$\ldots$$

$$U_{N!} := \{s : s = \pi_{N!}(x) \text{ for all } x \in U_{N!-1}\}.$$

This construction partitions the space without the singular collision points into $N!$ dynamically equivalent regions; there is no point in M which is not contained in some U_i and no two points in any one U_i are dynamically equivalent. Therefore, any one of these regions suffices to completely represent all the observationally (dynamically, in this case) distinct states of the system. That is, any one of these regions may be considered to be the sample space: $\Omega = U_i$ for any i. Suppose that this region is equipped with any measure of the form

$$\mu(U_i) := \int_{U_i} P(p, q)\, dp_1 ... dq_n = 1 \text{ for some } i \tag{5.6.1}$$

where $dq...dp = \omega^n$ is the symplectic volume form (see footnote 5).

If we now follow Gibbs and let P satisfy Equation (5.5.2) (so we let P be some function of the Hamiltonian), then μ will be preserved under those canonical transformations π_i which are also dynamical symmetries: ω^n is preserved by canonical transformations and P is preserved by dynamical symmetries. For the Hamiltonian we are considering, all of the permutations are dynamical symmetries, therefore

$$\mu(U_i) = 1 \text{ for all } i \tag{5.6.2}$$

from which, recalling that $M = \cup_{i=1}^{N!} U_i$, it follows that

$$\mu(M) = \sum_{i=1}^{N!} \mu(U_i) = N!\mu(U_i) = N!. \tag{5.6.3}$$

Therefore, when we want to compute the probability of the system being in region $R \subseteq M$ using Equation (5.2.5) we substitute in Equation (5.6.3) to get $\mathcal{P}(R) = \mu(R)/N!$. This concludes the derivation of the $N!$ factor and justifies why we must divide the measure of some region R by $N!$ to correctly compute the probability of that region. But I must emphasize that this procedure is a specific case of a more general and standard statistical procedure. If we want to compute the probability that the system lies in some region $R \subset M$, we first compute the measure of that region, $\mu(R)$ and then divide by $\mu(M)$ according to Equation (5.2.5). Usually it is implicitly assumed that $\mu(M) = 1$; it is assumed that the manifold is the sample space. One of the main results of this paper is that this is not always the case and this fact can be used to solve the $N!$ problem. In the general case where the sample space is not the whole manifold, $\mu(M) \neq 1$. In this particular example, $\mu(M) = N!$.

This argument in this Section generalizes: if the Hamiltonian satisfies $h(\pi(s)) = h(s)$ for all s and for all $\pi \in S_n$ where S_n is some permutation subgroup of S_N, then $\mu(M) = n!$. It may also be the case that the Hamiltonian satisfies $h(\pi(s)) = h(s)$ for all s and for all $\pi \in S_n$ and for all $\pi \in S_m$ where S_m is an altogether different subgroup. In this case $\mu(M) = n!m!$.

The $N!$ factor is often perceived as a ‘correction’, indicating that some reasoning over and above the standard statistical reasoning is needed to justify its inclusion. What I have shown is that this is not the case. The $N!$ arises perfectly naturally in the course of statistical reasoning as the normalisation factor $\mu(M)$ when all the permutations are dynamical symmetries. The reason it is perceived as a correction factor, is because division by $\mu(M) = 1$ renders itself invisible in cases when the manifold is the sample space. Thus, when the $N!$ factor appears, it looks like we have added something in that was not there before.

5.7 Conclusion

In this paper, I have given a solution to the $N!$ problem, based on the dynamical equivalence of the states and the dynamical indistinguishability of particles, whose precise definitions were inspired by considerations of observation, and its importance in specifying the sample space of any statistical theory.

These considerations of observation were key to my analysis. As we saw, the literature is most focused on whether two states are really 'one and the same' or whether particles are really 'identical'. This ontological view of the matter was replaced by the more epistemic notions of observational equivalence and observational indistinguishability respectively[22]. But why take this more 'epistemic' route with respect to the $N!$ problem in statistical mechanics? The first reason was that the basic object in a statistical theory is the probability space, one of whose elements being the sample space, typically characterized as containing all and only states which are observationally distinct from one another. Thus, epistemic notions enter into the very construction of the mathematical object we study

---

[22] I must emphasize at this point that the epistemic flavour of my analysis enters in the construction of the sample space, before any probability measures are defined on the space; the epistemic flavour does not enter into the interpretation of the probability measure. In fact, my analysis here is compatible with any interpretation of statistical mechanical probabilities: nothing in my argument requires us to make any choice of probability interpretation.

in statistical mechanics. The second reason was that the notion of particles sharing intrinsic properties like mass and charge was best interpreted formally and precisely in terms of observables; the state of the system only gives us position and momentum values, so in order to access their mass or charge, we must look to certain observables like the kinetic energy or the electrical potential, for example.

The notion of observational equivalence was further narrowed to the notion of dynamical equivalence, taking the set of observables with respect to which to judge the equivalence of states to be the singleton set containing the Hamiltonian. It is granted that dynamical equivalence is not the only sense in which states may be said to be observationally equivalent, and hence nor is dynamical indistinguishability the only sense in which particles may be said to be indistinguishable. This is accommodated in my analysis in the fact that the different senses of particle (in)distinguishability and state (in)equivalence correspond to different choices of the set $\{f_i\}$.

Although I have restricted my discussion to classical Hamiltonian mechanics, the characterisation of observational and dynamical equivalence I develop in this chapter will carry over straightforwardly to quantum theory; the characterisation of indistinguishable classical particles in terms of the permutation symmetry of the Hamiltonian takes the form of the permutation operator commuting with the Hamiltonian. The framework presented here therefore promises to make interesting and important connections between classical and quantum physics in the context of particle indistinguishability.

Chapter 6

Conclusion

6.1 Summary

This thesis has presented a philosophically novel, though mathematically well-known, formulation of thermodynamics and associated foundational consequences and given mathematically precise and philosophically grounded definitions of identity and indistinguishability in thermodynamics and classical statistical mechanics.

Chapter 2 demonstrated the philosophical utility of the geometric formulation of thermodynamics, grounding the distinction between heat and work in a distinction between hidden and observed degrees of freedom. This allowed us to infer a novel interpretation of the thermodynamic entropy as an aggregate variable representing the energy contributions from all the hidden degrees of freedom. I used the formulation to define a thermodynamic system as the pair (M, f) where M is the manifold representing the thermodynamic equilibrium states and f is the fundamental relation. This, in turn, allowed me to express the relationship between subsystems and their composite with mathematical precision. Building on this definition of composite system allowed a definition of thermodynamic equilibrium and equilibration in the geometric formalism. Combining this with the work of Wightman resulted in a helpful clarification of the second law of thermodynamics.

Having established the thermodynamic formalism, I applied it to the philosophical problem of how to best represent mixtures in Chapter 3. Using two plausible representational principles, I concluded that Gibbs' representation is the best model, despite the standard representation being more commonly used. One consequence of this is a reassessment of the status of 'partial pressures' in thermodynamics. In the context of the Gibbs representation they are nothing more than a mathematical fiction and result from a symbolic confusion between a variable and the value of a variable. By adopting Gibbs' representation, we are also forced to rethink the representation of the mixing process. Using the definitions of equilibration and composite system from Chapter 2, I proposed a representation of the mixing process which makes no use of semipermeable membranes.

With the representation of mixture in hand, I tackled identity and Gibbs' paradox in Chapter 4. I argued that identity, or indeed any purportedly thermodynamic concept, must be defined without appeal to extra-theoretic ideas if it is to be used for quantitative or foundational judgements. With this is mind, I proposed a definition of identity based on the homogeneity of the fundamental relation and the definition of composite system. This made sense of Gibbs' paradox by giving clear and precise definitions of identity and difference.

We finally turned to the $N!$ problem, the statistical mechanical version of the Gibbs paradox. I proposed a new general interpretation of statistical theories, namely that the sample space on which the probability measure is defined is constrained by observables. This, together with my definition of observational indistinguishability, allowed the phase space of the mechanical system to be partitioned into a number of observationally indistinguishable regions, each of which may be considered to be a sample space. In the special case where we judge indistinguishability of states with respect to the free particle hamiltonian, the $N!$ factor is rigorously derived and should no longer be seen as a 'correction'.

This thesis has been about so much more than just Gibbs' paradox. Just like other paradoxes in the history of philosophy, it was worth thinking about, not just for the pleasure of solving a (admittedly niche) problem in the foundations of physics, but because it points to deep, foundational issues which merit investigation. The foundational items that have been highlighted and clarified on our journey to this solution are numerous: the importance of representing the identity relation in the thermodynamic formalism; a definition of a thermodynamic model was proposed in geometric terms with the fundamental relation taking centre stage; the link between the homogeneity of thermodynamic models and the scaling symmetry of the fundamental relation played a key role in the definition of the scaling criterion, and the notions of subsystem and composite system received a precise and general analysis. Gibbs' paradox has turned out to be surprisingly fruitful.

The philosophical implications of the proposed criterion of identity for thermodynamic models are particularly interesting. A plausible criterion of identity in terms of isomorphism of the systems' formal structures turned out to lead to unwelcome conclusions. The strategy I developed was to define the 'difference' between the systems as the failure of an identity relation that holds between subsystems of a composite rather than between systems themselves; if two thermodynamic models are not subsystems of some composite, then they cannot be relata of the identity relation specified by the scaling criterion.

The traditional conception of identity which appeals to the intrinsic physical and chemical properties of the atomic constituents of the thermodynamic models is purely heuristic since the properties of atoms are not capable of representation in macroscopic thermodynamics. But there is a stronger reason we should not appeal to such concepts: thermodynamics is an extremely powerful and general theory for when we do not have access to the underlying microphysics of physical systems. As it happens, in the case of gases, we do know about the physics of the micro-constituents and this may be why we are naturally attracted to expressing identity in terms of the properties of those constituents. But, in general, we should not assume that we have access to the microphysics and so we should not let definitions of concepts such as identity to be tied to it. This is why I strived to explicate identity of thermodynamic models in purely thermodynamic terms.

Thermodynamics, expressed in the geometric formulation presented here, is an extremely powerful and general theory for modelling and predicting the behaviour of physical systems when the underlying microphysics is unknown, ignored or inaccessible. The full philosophical implications of this formulation have yet to be fully explored but, if this analysis is anything to go by, is sure to be fruitful.

6.2 Open Questions

Many open questions about Gibbs' paradox, identity and indistinguishability, and the foundations of thermodynamics and statistical mechanics remain unanswered or are new questions raised in this work.

6.2.1 Connections between the geometric formulation and other formulations

In Chapter 2, I focused on comparing orthodox thermodynamics and geometric thermodynamics. I delineated some differences and similarities between the two formulations, but the task of fully comparing them is far from complete. One aspect of this comparison which has been left largely untouched is the notion of 'process' in thermodynamics. Processes, particularly cyclic processes, seem to play a foundational role in orthodox thermodynamics, serving to ground the definitions of certain functions of state. By contrast, the notion of process in the geometric formulation rarely, if ever, appears. This is why Gibbs' formulation of thermodynamics, from which the contact formulation is inspired, is sometimes called *thermostatics*.

Understanding this aspect of the comparison is extremely important for our understanding of the second law, since discussions of the orthodox version of the law make unavoidable mention of 'reversible' and 'irreversible' processes, whereas the geometric version of the law makes no such mention. Why do processes play such a foundational role in orthodox thermodynamics and not in thermostatics?

The geometric formulation of thermodynamics is not the only thermodynamic theory claiming to be mathematically rigorous. Lieb and Yngvason (1999) proposed an axiomatic formulation, inspired by the work of Carathéodery, which has received widespread philosophical attention. How do the second laws of the two formulations compare? Is there a foundational difference between them in what quantities they take to be fundamental?

6.2.2 How to reduce the geometric second law

At some point, we must face the elephant in the room: reduction. The reduction of thermodynamics to statistical mechanics historically has dominated philosophical discussions of thermal physics. It is an extremely difficult problem partly, I speculate, because we not really got clear yet on the mathematical and physical structure of thermodynamics and statistical mechanics.

Now that we have another philosophically fruitful formulation of thermodynamics, we must consider the inevitable question: can the laws of this formulation be reduced in some sense to the laws of statistical mechanics?

6.2.3 The $N!$ problem in Boltzmannian statistical mechanics

My discussion of the $N!$ problem in Chapter 5 is set in Gibbsian statistical mechanics. This is done for good reason, since comments in the book by Gibbs in which he first laid out the theory precipitated interest in the $N!$ problem.

However, a quirk of the history of Gibbs' paradox is that many modern formulations of the $N!$ problem are set in Boltzmannian statistical mechanics, with focus being centred on the apparent non-extensivity of the entropy. Does setting the problem in Boltzmannian statistical mechanics give us any new insight? Why did Boltzmann not arrive at something like the $N!$ problem? Can Boltzmannian statistical mechanics be formulated in a similarly coherent fashion to Gibbsian statistical mechanics?

6.2.4 How to define quantum particle indistinguishability

The framework I presented in Chapter 5 brings discussion of classical particle indistinguishability closer to discussions of quantum particle indistinguishability. Our basic conclusion was that the indistinguishability of classical particles can be grounded in the permutation symmetry of a set of classical observables (most typically the hamiltonian). This idea carries over naturally into quantum physics: when we want to say that particles are indistinguishable, we state that the probability amplitudes must be invariant under exchange of particle labels (i.e. acting on Fock space with the permutation operator leaves the amplitude the same) and that the hamiltonian is similarly invariant. Further work would involve carrying out this idea in detail and investigating the relationship of this kind of indistinguishability with the other characterisations in terms of the symmetrisation postulate and the uncertainty principle mentioned in the introduction.

Appendix A

Appendix

A.1 Deriving Equations (VK1), (VK2) and (VK3)

A.1.1 Equation (VK1)

Throughout van Kampen's derivation of Equation (4.2.3) (which provides the base block for deriving Equations (VK1)–(VK3)), he treats N as a constant. We can see this explicitly in his statement of the First Law, Equation (4.2.1), in which the only variables are U, S, V, T and p. It is also implicit in his derivation since, although he does not write it down explicitly, it is necessary to assume that $dU = (3/2)NRdT$ follows from $U = (3/2)NRT$ in order to derive Equation (4.2.3). It is again implicit in the derivation of Equation (4.2.3) where C absorbs terms containing N (this happens in the step from Equation (3.2.5) to Equation (4.2.3)).

But he also seems to want to treat N as a variable, for, towards the beginning of his article, he claims that the Gibbs paradox "is a consequence of the way in which the entropy depends on the number of molecules, and I therefore start out in section 2 by sketching how this dependence is established in thermodynamics." (van Kampen, 1984, p. 304). He is therefore clearly engaged in the business of finding entropy as a function of N which is clearly treating N as a variable.

Therefore we can see that he begins his analysis apparently stating forcefully that it is a variable and, indeed, that the whole Gibbs paradox rests on this functional dependence, and yet he carries out his derivation of Equation (4.2.3) assuming it is a constant, before using his conventions to turn it back into a variable again. Flipping between these interpretations of N is clearly being inconsistent; it cannot be both a constant and a variable. So, which is it? From the point of view of giving a rigorous mathematical derivation of the thermodynamic entropy for the ideal gas, this is a problem.

In this Section, I derive (VK1) by assuming that N is a variable. We begin by assuming that the fundamental relation of the ideal gas takes the form

$$S = f(U, V, N) \tag{A.1.1}$$

Taking the derivative, we see that this takes on the familiar form of the first law of thermodynamics:

$$TdS = dU + pdV - \mu dN \tag{A.1.2}$$

This fundamental relation is assumed to be first-order homogeneous and so it follows from Euler's theorem on homogeneous functions that

$$S = \frac{1}{T}U + \frac{p}{T}V - \frac{\mu}{T}N \tag{A.1.3}$$

We know the following two equations of state from experiment:

$$1/T = 3NR/2U; \qquad p/T = NRV. \tag{A.1.4}$$

We may derive an expression for μ/T from these two because the intensive parameters of a first-order homogeneous relation are not independent. We may show this as follows.

Taking the derivative of Equation (A.1.3), we have:

$$dS = Ud\left(\frac{1}{T}\right) + \frac{1}{T}dU + Vd\left(\frac{p}{T}\right) + \frac{p}{T}dV - Nd\left(\frac{\mu}{T}\right) - \frac{\mu}{T}dN \tag{A.1.5}$$

Subtracting from this Equation (A.1.2) we get what is known as the *Gibbs-Duhem relation*:

$$0 = Ud\left(\frac{1}{T}\right) + Vd\left(\frac{p}{T}\right) - Nd\left(\frac{\mu}{T}\right) \tag{A.1.6}$$

This equation shows that the intensive parameters are not independent, since it may be integrated to find the third equation of state from the other two. Taking the derivative of Equations (A.1.4) and substituting into Equation (A.1.5) we get

$$Nd\left(\frac{\mu}{T}\right) = U\left(\frac{3R}{2U}dN - \frac{3NR}{2U^2}dU\right) + V\left(\frac{R}{V}dN - \frac{NR}{V^2}dV\right) \tag{A.1.7}$$

from which it follows that

$$d\left(\frac{\mu}{T}\right) = \frac{5R}{2N}dN - \frac{3R}{2U}dU - \frac{R}{V}dU. \tag{A.1.8}$$

We may now integrate this to get

$$\frac{\mu}{T} = \frac{5}{2}R\ln N - \frac{3}{2}R\ln U - R\ln V + K = R\ln\left[\frac{N^{5/2}}{U^{3/2}V}\right] + K \tag{A.1.9}$$

where K is the constant of integration.

Substituting the equations of state (A.1.4) and (A.1.9) into (A.1.3) we get the fundamental relation of the ideal gas as a function of U, V, N. It is easy to check that simplifying this equation and eliminating V and U in favour of p and T using the equations of state, we arrive at Equation (VK1).

A.1.2 **Equation** (VK2)

In this Appendix, I give detailed physical justifications for the definitions of the variables for (M_G, f_G) which I claim represents the ideal gas mixture.

- S is defined by the additivity postulate to be:

  $$S := \sum_{i}^{k} S^{(i)} = \sum_{i=1}^{k} f^{(i)}(U^{(i)}, V^{(i)}, N^{(i)}) \tag{A.1.10}$$

  where each of the $f^{(i)}$ is the fundamental relation of a single ideal gas as given in Equation (3.2.9).

- U is defined to be:

  $$U := \sum_{i=1}^{k} U^{(i)}. \tag{A.1.11}$$

  This definition expresses the physical assumption that the energy of the composite system is the sum of the energies of the individual subsystems.

This is a typical physical assumption for the ideal gas mixture which is often motivated by a 'billiard ball' model of an ideal gas in which there are no interactions between the individual molecules. We also impose the condition

$$\partial f^{(1)}/\partial U^{(1)} = ... = \partial f^{(k)}/\partial U^{(k)} \tag{A.1.12}$$

which expresses the assumption for the simple ideal gas mixture that the temperatures of all the subsystems are the same. This long series of equalities is $U^{(1)}/N^{(1)} = ... = U^{(k)}/N^{(k)}$ where it can be shown that $U^{(i)}/N^{(i)} = U/\sum_i N^{(i)}$ for all i. This assumption does not form part of the definition of the variable U; it is an additional assumption made to incorporate a physical fact about the type of system we are dealing with.

- Variable V is defined to be:

$$V := V^{(1)} = ... = V^{(k)} \tag{A.1.13}$$

This definition states that the volumes of the subsystems are all equal and that the volume of the mixture V is equal to any one of those. This incorporates part of the definition of mixture which says that the components all occupy the very same volume and, therefore, that the volumes of the components are identical. We also impose the condition

$$\partial f/\partial V := \sum_{i=1}^{k} \partial f^{(i)}/\partial V^{(i)} \tag{A.1.14}$$

which, together with Equation (A.1.12), expresses the assumption that the pressure of the mixture is equal to the sum of the pressures of the individual components. This is more familiarly known as Dalton's Law.

- The variables N_i are defined to be:

  $$N_i := N^{(i)} \text{ for } i = 1, ..., k \tag{A.1.15}$$

  This definition states that the mole number of component i in the mixture is equal to the mole number of subsystem (i). It also captures precisely the intuition that components of a mixture are subsystems of the mixture.

With these definitions and assumptions the fundamental relation of the ideal gas mixture becomes

$$S = f_*(U, V, N_1, ..., N_k) = \frac{1}{T}U + \frac{p}{T}V - \sum_i \frac{\mu_i}{T} N_i \tag{A.1.16}$$

where

$$\frac{1}{T} = \sum_i \frac{3}{2} R \frac{N_i}{U} \tag{A.1.17}$$

$$\frac{p}{T} = \sum_i R \frac{N_i}{V} \tag{A.1.18}$$

$$\frac{\mu_i}{T} = R \ln \left[\frac{N_i \left(\sum_j N_j \right)^{3/2}}{V U^{3/2}} \right] + K_i \text{ for } i = 1, ..., k \tag{A.1.19}$$

From this explicit fundamental equation, it is easy to see that the pair (M_G, f_G) is a thermodynamic model.

By eliminating U in favour of T, defining the 'partial pressures' $p_i := N_i/V =$

$pN_i/(\sum_i N_i)$, and absorbing all constants into the K_i, we recover van Kampen's Equation (VK2) for k components in the following form:

$$S = \frac{5}{2}\left(\sum_{i=1}^{k} N_i\right) R\ln T - \sum_{i=1}^{k}(N_i R\ln p_i) + \left(\sum_{i=1}^{k} N_i K_i\right) \tag{VK2$'$}$$

where the summation indices run over the number of components in the mixture.

A.1.3 Equation (VK3)

The variables for $(M^{\text{VK3}}, f^{\text{VK3}})$ are defined as follows.

- S is defined, as before, by the additivity postulate:

$$S := \sum_{i}^{k} S^{(i)} = \sum_{i=1}^{k} f^{(i)}(U^{(i)}, V^{(i)}, N^{(i)}). \tag{A.1.10}$$

- As before, U is defined to be:

$$U := \sum_{i=1}^{k} U^{(i)}. \tag{A.1.11}$$

  with

$$\partial f^{(1)}/\partial U^{(1)} = ... = \partial f^{(k)}/\partial U^{(k)} \tag{A.1.12}$$

  which corresponds to equality of temperatures. Again, this long series of equalities is $U^{(1)}/N^{(1)} = ... = U^{(k)}/N^{(k)}$ where it can be shown that $U^{(i)}/N^{(i)} = U/\sum_i N^{(i)}$ for all i. It follows that $U^{(i)} = UN^{(i)}/\sum_i N^{(i)}$.

- V is defined to be:

$$V := \sum_{i}^{k} V^{(i)} \tag{A.1.20}$$

This definition states that the volume of the whole is the sum of the volumes of the subsystems and is thus quite different from the definition expressed in (A.1.13). We further impose the condition

$$\partial f^{(1)}/\partial V^{(1)} = ... = \partial f^{(k)}/\partial V^{(k)} \tag{A.1.21}$$

which, together with (A.1.12), is the assumption of mechanical equilibrium, i.e. that the pressures of the subsystems are equal. The long series of equalities is $N^{(1)}/V^{(1)} = ... = N^{(k)}/V^{(k)}$ where it can be shown that $V^{(i)} = VN^{(i)}/\sum_i N^{(i)}$.

- N is defined to be:

$$N := \sum_{i=1}^{k} N^{(i)}. \tag{A.1.22}$$

This assumption states that the mole number of the composite system is the sum of the mole numbers of the subsystems. With this definition, we are no longer considering simple mixtures, which always have more than one configuration variable interpreted as mole number. Here, the composite system has one configuration variable interpreted as mole number and so it is not a mixture.

It follows from the Gibbs-Duhem relation that the intensive variables temperature, pressure and chemical potential are not independent. Therefore, from the previous requirements of equality of temperatures and pressures, it follows that the chemical potentials will be equal. Hence, we have that

$$\partial f^{(1)}/\partial N^{(1)} = ... = \partial f^{(k)}/\partial N^{(k)} \tag{A.1.23}$$

which, together with (A.1.12), is the assumption of chemical equilibrium, meaning that there is no transfer of matter from one subsystem to another.

Applying these definitions, we arrive at the fundamental relation

$$S = \frac{1}{T}U + \frac{p}{T}V - \frac{\mu}{T}N \tag{A.1.24}$$

with the equations of state

$$\frac{1}{T} = \frac{3}{2}R\frac{N}{U}; \tag{A.1.25}$$

$$\frac{p}{T} = R\frac{N}{V}; \tag{A.1.26}$$

$$\frac{\mu}{T} = R\ln\left[\frac{N^{5/2}}{VU^{3/2}}\right] + K. \tag{A.1.27}$$

If we eliminate U in favour of T and V in favour of p, and define $p^{(i)} := N^{(i)}RT/V$ while also writing $N = \sum_i N^{(i)}$ and $p = \sum_i p^{(i)}$ we recover van Kampen's (VK3) in the following form:

$$S = \frac{5}{2}R\left(\sum_i^k N^{(i)}\right)\ln T - \left(\sum_i^k N^{(i)}\right)R\ln\left[\sum_i^k p^{(i)}\right] + \left(\sum_i^k N^{(i)}\right)K. \tag{VK3$'$}$$

A.2 Gibbs' Representation

In this Appendix, we show some of the consequences of Equation (3.3.7) and show that it is the full Legendre transform of the explicit form of $f_G(U, V, N_1, \ldots, N_k)$.

Let (M, f) be a thermodynamic system with variables $X_0, ..., X_k$ and first-order homogeneous fundamental relation between the extensive variables

$$X_0 = f(X_1, ..., X_k). \tag{A.2.1}$$

Let us assume that we have an explicit form for this relation. Then we can derive explicit expressions for the intensive parameters, which we denote

$$F_i := \partial f / \partial X_i \text{ for } i = 1, .., k \tag{A.2.2}$$

as functions of $X_1, ..., X_k$. Let us also consider the expression

$$g(F_1, ..., F_k) := X_0 - \sum_{i=1}^{k} F_i X_i \tag{A.2.3}$$

which is the full Legendre transform of the fundamental relation f. Eliminating X_0 from this expression using Equation (A.2.1) and then eliminating the X_i using Equation (A.2.2), we arrive at a relation between all the intensive parameters. In fact, because Equation (A.2.1) is first-order homogeneous, by Euler's theorem it can be written as

$$X_0 = \sum_{i=1}^{k} X_i F_i. \tag{A.2.4}$$

Substituting Equation (A.2.4) into (A.2.3) it follows that

$$g(F_1, ..., F_k) = 0 \tag{A.2.5}$$

which is the desired explicit relation between the intensive parameters we have been seeking.

The converse is also true: we can find the explicit relation between the extensive parameters assuming we have an explicit relation between the intensive parameters. Therefore, let us assume we have such an explicit relation $g(F_1, ..., F_k) = 0$. X_0 is given by Equation (A.2.4), so to find X_0 as a function of $X_1, ..., X_k$, we must find each F_i as a function of $X_1, ..., X_k$. To do this, note that the derivative of Equation (A.2.3) is $dg = dX_0 - \sum_{i=1}^{k} (X_i dF_i + F_i dX_i)$, and note that it follows from Equation (A.2.1) that $dX_0 = \sum_{i=1}^{k} F_i dX_i$ which means that $dg = -\sum_i X_i dF_i$. From this we deduce

$$-X_i = \frac{\partial g}{\partial F_i} \text{ for } i = 1, ..., k \tag{A.2.6}$$

which we can then use to eliminate $F_1, ..., F_k$ in favour of $X_1, ..., X_k$ in Equation (A.2.4), yielding an explicit relation $X_0 = f(X_1, ..., X_k)$ between the extensive variables.

With this clarification, we can better understand and reconstruct Gibbs' reasoning in which he derives certain thermodynamic properties from Equation (3.3.7) and, even though he does not explicitly do it, I will show how to derive the extensive variable fundamental relation of the ideal gas mixture from Equation (3.3.7).

The differential of Equation (3.3.7) is

$$dp = \frac{\partial p}{\partial T} dT + \sum_i^n \frac{\partial p}{\partial \mu_i} d\mu_i \tag{A.2.7}$$

The Gibbs-Duhem relation derived from the fundamental relation in the energy representation for the thermodynamic system (M_G, f_G) is

$$0 = SdT - Vdp + \sum_i^n N_i d\mu_i \tag{A.2.8}$$

which rearranges to

$$dp = \frac{S}{V} dT + \sum_i^n \frac{N_i}{V} d\mu_i \tag{A.2.9}$$

Comparing (A.2.9) and (A.2.7) and using the expression for p in Equation (3.3.7) we derive that

$$S/V = \sum_i^n \left(C_i T^{3/2} e^{\mu_i/T} \left(\frac{5}{2} - \frac{\mu_i}{T} \right) \right) \tag{A.2.10}$$

$$N_i/V = C_i T^{3/2} e^{\mu_i/T} \tag{A.2.11}$$

Substituting (A.2.11) into (3.3.7) and (A.2.11) into (A.2.10) Gibbs obtains Equations (3.3.8) and (3.3.9).

A.3 Calculating the Entropy of Mixing

Let us begin by understanding the traditional representation of mixing based on the representation of mixture reconstructed. To represent the mixing equilibration process, we view (M_s, f_s) as being composed out of subsystems $(M_s^{(1)}, f_s^{(1)})$ and

$(M_s^{(2)}, f_s^{(2)})$ in the way outlined in the definition of K-equilibrium. Mixing is the transition from the unmixed state (where the values of the subsystems' extensive variables do not extremise the entropy) to the mixed state (where they do) as a result of the removal of a constraint on the flow of the extensive quantities.

In the context of Gibbs' paradox, the temperatures of the samples each side of the partition are equal, leaving only the volume and mole number constraints to be removed. We can think of mixing as removing either the volume constraint or the mole number constraint, but removing one of them will cause both remaining intensive parameters to equalise. The transition from the initial to the final state as a result of the removal of the volume constraint is summarised in Table A.1.

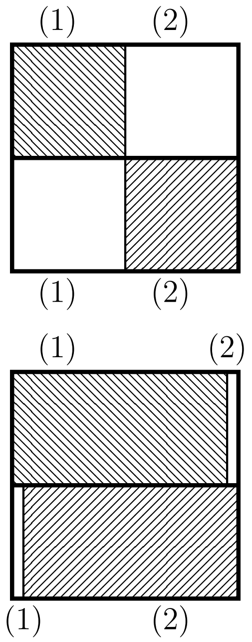


Figure A.1: Depiction of mixing represented as two free expansions. In order for this to represent a mixture, we are to imagine that V_1 and V_2 have the same value and ‘overlap’.

Variable	Initial	Final
$U^{(1)}$	$U/2$	$U/2$
$U^{(2)}$	$U/2$	$U/2$
$V_1^{(1)}$	$V_1/2$	V_1
$V_1^{(2)}$	$V_1/2$	0
$V_2^{(1)}$	$V_2/2$	0
$V_2^{(2)}$	$V_2/2$	V_2
$N_1^{(1)}$	N_1	N_1
$N_1^{(2)}$	0	0
$N_2^{(1)}$	0	0
$N_2^{(2)}$	N_2	N_2

Table A.1: Transition from initial to final state by removal of the volume constraint, when representing mixing on the basis of (M_s, f_s).

This final state may be calculated along the lines explained in Section **??**. Having decomposed the system into subsystems, we allow the volume to ‘flow’ between the subsystems, reaching values of $V_1^{(1)}$, $V_1^{(2)}$, $V_2^{(1)}$ and $V_2^{(2)}$ which extremise the entropy:

$$0 = dS = \left(\frac{p_1^{(1)}}{T} - \frac{p_1^{(2)}}{T} \right) dV_1^{(1)} + \left(\frac{p_2^{(1)}}{T} - \frac{p_2^{(2)}}{T} \right) dV_2^{(1)}, \tag{A.3.1}$$

concluding that the condition for equilibrium is $p_1^{(1)}/T = p_1^{(2)}/T$ and $p_2^{(1)}/T = p_2^{(2)}/T$. Imposing this condition implies that the gases expand to fill the volume of the container, with final values[1] as illustrated in Table A.1.

We may calculate the entropy change by calculating the difference between the initial and final value of $S_s = S_s^{(1)} + S_s^{(2)}$. Since it is the volume flowing between the subsystems that generates this entropy increase and there is no flow of energy or mole number between the subsystems, we find that the entropy change is the difference between the initial and final values of the expression

$$R\left(N_1^{(1)} \ln V_1^{(1)} + N_1^{(2)} \ln V_1^{(2)} + N_2^{(1)} \ln V_2^{(1)} + N_2^{(2)} \ln V_2^{(2)}\right) \tag{A.3.2}$$

where we have used the fundamental relation f_s in Equation (3.3.1). Substituting in the initial and final values of these variables taken from Table A.1, we can calculate that the entropy change is $2(N_1 + N_2)R \ln 2$.

We now turn to the representation of mixing based on (M_G, f_G). In this case, the mixing process is represented differently, due to the fact that the system has different extensive variables and hence different kinds of internal constraints on the flow of extensive quantities.

---

[1] These values are calculated as follows. Substituting in the equation of state $pV = NRT$ for the ideal gas into the conditions of equilibrium $p_i^{(1)}/T = p_i^{(2)}/T$ we get $N_i^{(1)}/V_i^{(1)} = N_i^{(2)}/V_i^{(2)}$. Given that the mole numbers of the subsystems remain constant and the volumes are subject to $V_i = V_i^{(1)} + V_i^{(2)}$, we get $N_i^{(1)}/V_i^{(1)} = N_i^{(2)}/V_i^{(2)} = N_i/V_i$. Rearranging to make V_i^j for $i, j = 1, 2$ the subject, we calculate the values given in Table A.1.

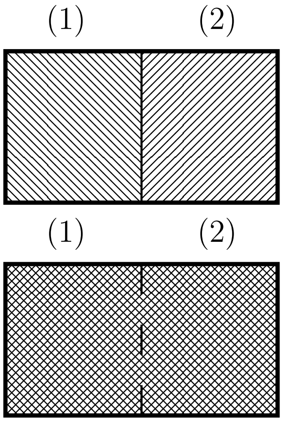


Figure A.2: Depiction of the mixing process according to the Gibbs representation. The initial, unmixed system is depicted on top and the final, mixed system on the bottom.

Variable	Initial	Final
$U^{(1)}$	$U/2$	$U/2$
$U^{(2)}$	$U/2$	$U/2$
$V^{(1)}$	$V/2$	$V/2$
$V^{(2)}$	$V/2$	$V/2$
$N_1^{(1)}$	N	$N/2$
$N_1^{(2)}$	0	$N/2$
$N_2^{(1)}$	0	$N/2$
$N_2^{(2)}$	N	$N/2$

Table A.2: Table illustrating the initial and final values of the extensive variables of the subsystems, before and after the removal of the constraint on the mole number.

We assume that the pressures and temperatures of each subsystem are equal by removing constraints on the flow of energy and volume. However, we assume the system still has internal constraints on the flow of the mole numbers of the components in place. If we remove the constraint on mole number, the conditions for chemical equilibrium for each component will be satisfied: $\mu_1^{(1)}/T = \mu_1^{(2)}/T$ and $\mu_2^{(1)}/T = \mu_2^{(2)}/T$ as may be easily verified by running a similar calculation to Equation (A.3.1) to extremise the entropy. From these conditions and the initial values given in Table A.2, the final values of $N_1^{(1)}$, $N_1^{(2)}$, $N_2^{(1)}$ and $N_2^{(2)}$ may be calculated. The entropy change is obtained by calculating the initial and final values of $S_G = S_*^{(1)} + S_G^{(2)}$ which, using the fundamental relation f_G in Equation (A.1.16), is the difference between the initial and final values of the expression

$$R\left(N_1^{(1)}\ln N_1^{(1)}+N_1^{(2)}\ln N_1^{(2)}+N_2^{(1)}\ln N_2^{(1)}+N_2^{(2)}\ln N_2^{(2)}\right) \tag{A.3.3}$$

where we substitute in the values of the mole numbers from Table A.2. We find $\Delta S_G = 2(N_1 + N_2)R\ln 2$.

Comparing the representations of mixing, we see that they represent the process quite differently. In the process based on (M_s, f_s), mixing is initiated by removing the volume constraint, while in the process based on (M_G, f_G), mixing is initiated by removing the mole number constraint. As a result, the calculation of the entropy change is entirely different (the first relies on a flow of volume between the subsystems, while the second relies on the flow of mole number). Both calculations happen to give $2(N_1 + N_2)R\ln 2$. It is interesting to note that one arrives at the very same numerical answer for the entropy increase on mixing despite the systems and processes being entirely different. It is not a coincidence, however, since it follows from Gibbs' theorem (see Section 4.2.2) that the entropy of a mixture of ideal gases is the sum of the entropies each component would have if they occupied a container with the same volume and temperature.

Bibliography

Adkins, C. J. (1983). *Equilibrium thermodynamics.* Cambridge University Press.

Arnol'd, V. I. (1978). *Mathematical Methods of Classical Mechanics.* Trans. by K. Vogtmann and A. Weinstein. Graduate Texts in Mathematics. Springer New York.

Arnold, V. I. (1990). "Contact Geometry: The Geometrical Method of Gibbs's Thermodynamics". In: *Proceedings of the Gibbs Symposium: Yale University, May 15-17, 1989.* Ed. by D.G. Caldi and G.D. Mostow. American Mathematical Society, pp. 163–180.

Barsky, Eugene (2017). *Gibbs' Entropic Paradox and Problems of Separation Processes.* Elsevier.

Belot, Gordon (2003). "Notes on Symmetries". In: *Symmetries in Physics: Philosophical Reflections.* Ed. by Katherine A. Brading and Elena Castellani. Cambridge University Press, pp. 393–412.

— (2018). "Fifty Million Elvis Fans Can't be Wrong". In: *Noûs*, pp. 946–981.

Billingsley, P. (1995). *Probability and Measure.* Wiley.

Blundell, Stephen J. and Katherine M. Blundell (2010). *Concepts in Thermal Physics.* 2nd ed. Oxford University Press.

Bridgman, Percy W. (1961). *The Nature of Thermodynamics.* New York: Harper & Brothers.

Buchdahl, H.A. (1966). *The Concepts of Classical Thermodynamics.* Cambridge Monographs on Physics. Cambridge University Press.

Burke, W.L. (1985). *Applied Differential Geometry.* Cambridge University Press.

Butterfield, Jeremy (2007). "On Symplectic Reduction in Classical Mechanics". In: *Philosophy of Physics.* Ed. by Jeremy Butterfield and John Earman. Handbook of the Philosophy of Science. Amsterdam: North-Holland, pp. 1–132.

Callen, H.B. (1960). *Thermodynamics.* J. Wiley.

Cardwell, D.S.L. (1971). *From Watt to Clausius: The Rise of Thermodynamics in the Early Industrial Age.* Cornell University Press.

Carnot, S. (1872/1890). *Reflections on the Motive Power of Heat and on Machines Fitted to Develop that Power.* Landmarks of science. J. Wiley.

Castellani, Elena (2003). "Symmetry and Equivalence". In: *Symmetries in Physics: Philosophical Reflections.* Ed. by Katherine Brading and Elena Castellani. Cambridge University Press, pp. 425–436.

Caulton, Adam and Jeremy Butterfield (Aug. 2011). "On Kinds of Indiscernibility in Logic and Metaphysics". In: *The British Journal for the Philosophy of Science* 63.1, pp. 27–84.

— (2012). "Symmetries and Paraparticles as a Motivation for Structuralism". In: *The British Journal for the Philosophy of Science* 63.2, pp. 233–285.

Chang, Hasok (2003). "Preservative Realism and its Discontents: Revisiting Caloric". In: *Philosophy of Science* 70.5, pp. 902–912.

Clausius, R. (1851). “On the moving force of heat, and the laws regarding the nature of heat itself which are deducible therefrom”. In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2.8, pp. 1–21.

— (1879). *The Mechanical Theory of Heat.* Trans. by Walter R. Browne. Macmillan & Co.

Darrigol, Olivier (1991). “Statistics and Combinatorics in Early Quantum Theory, II: Early Symptoma of Indistinguishability and Holism”. In: *Historical Studies in the Physical and Biological Sciences* 21 (2), pp. 237–298.

Denbigh, K. G. and J. S. Denbigh (1985). *Entropy in Relation to Incomplete Knowledge.* Cambridge University Press.

Denbigh, K. G. and M. L. G. Redhead (1989). “Gibbs’ Paradox and Non-Uniform Convergence”. In: *Synthese* 81.3, pp. 283–312.

Dieks, Dennis (1990). “Quantum Statistics, Identical Particles and Correlations”. In: *Synthese* 82.1, pp. 127–155.

— (2011). “The Gibbs Paradox Revisited”. In: *Explanation, Prediction, and Confirmation.* Ed. by Dennis Dieks et al. Dordrecht: Springer Netherlands, pp. 367–377.

— (2013). “Is There a Unique Physical Entropy? Micro versus Macro”. In: *New Challenges to Philosophy of Science.* Ed. by H. Andersen et al. Springer Verlag, pp. 23–34.

— (2014). “The Logic of Identity: Distinguishability and Indistinguishability in Classical and Quantum Physics”. In: *Foundations of Physics* 44 (12), pp. 1302–1316.

— (2018). “The Gibbs Paradox and Particle Individuality”. In: *Entropy* 20.6.

Dugdale, J. S. (1966). *Entropy and Low Temperature Physics.* Hutchinson & Co.

Elwell, D. and A. J. Pointon (1972). *Classical thermodynamics.* Penguin.

Fay, James A. (1965). *Molecular Thermodynamics.* Addison-Wesley.

Fletcher, Samuel C (2020). "On Representational Capacities, with an Application to General Relativity". In: *Foundations of Physics* 50.4, pp. 228–249.

Fong, Peter (1963). *Foundations of Thermodynamics.* Oxford University Press.

French, S. and D Krause (2006). *Identity in Physics: A Historical, Philosophical, and Formal Analysis.* OUP: Clarendon Press.

French, Steven (1989). "Identity and Individuality in Classical and Quantum Physics". In: *Australasian Journal of Philosophy* 67.4, pp. 432–446.

— (2000). "Identity and Individuality in Quantum Theory". In: *The Stanford Encyclopedia of Philosophy.* Ed. by Edward N. Zalta. Winter 2019. Metaphysics Research Lab, Stanford University.

French, Steven and Dean Rickles (2003). "Understanding Permutation Symmetry". In: *Symmetries in Physics: Philosophical Reflections.* Ed. by Katherine Brading and Elena Castellani. Cambridge University Press, pp. 212–238.

Frigg, Roman (2008). "A Field Guide to Recent Work on the Foundations of Statistical Mechanics". In: *The Ashgate Companion to Contemporary Philosophy of Physics.* Ed. by Dean Rickles. London: Ashgate, pp. 99–196.

Frigg, Roman and James Nguyen (2020). *Modelling nature: An Opinionated Introduction to Scientific Representation.* Springer.

Gibbs, J. W. (1873a). "A Method of Geometrical Representation of the Thermodynamic Properties of Substances by Means of Surfaces". In: *Transactions of the Connecticut Academy of Arts and Sciences.*

— (1873b). "Graphical Methods in the Thermodynamics of Fluids". In: *Transactions of the Connecticut Academy of Arts and Sciences.*

— (1878). “On the Equilibrium of Heterogeneous Substances”. In: *Transactions of the Connecticut Academy of Arts and Sciences* (3), 108–248 and 343–524.

— (1902). *Elementary Principles in Statistical Mechanics.* New Haven: Yale University Press.

Glazer, Mike and Justin S. Wark (2001). *Statistical Mechanics: A Survival Guide.* OUP.

Halvorson, Hans (2019). *The Logic in Philosophy of Science.* Cambridge University Press.

Hermann, R. (1973). *Geometry, Physics, and Systems.* Lecture Notes in Pure and Applied Mathematics. M. Dekker.

Hill, T.L. (1956). *Statistical Mechanics: Principles and Selected Applications.* Dover Books on Physics. McGraw-Hill.

Huang, Kerson (1963). *Statistical Mechanics.* John Wiley & Sons.

Jaynes, E. T. (1992). “The Gibbs Paradox”. In: *Maximum-entropy and Bayesian methods.* Ed. by G. Erickson, P. P. Neudorfer, and C. R. Smith. Dordrecht: Kluwer, pp. 1–21.

Kestin, J. (1971). *A Course In Statistical Thermodynamics.*

Klein, Martin J. (1958). “Note on a Problem Concerning the Gibbs Paradox”. In: *American Journal of Physics* 26.2, pp. 80–81.

Kondepudi, D. and I. Prigogine (1998). *Modern Thermodynamics From Heat Engines to Dissipative Structures.* John Wiley & Sons.

Kubo, Ryogo (1968). *Thermodynamics.* North-Holland.

Landsberg, Peter T. (1978). *Thermodynamics and Statistical Mechanics.* Oxford University Press.

Lavenda, Bernard (1991). *Statistical Physics: A Probabilistic Approach.* John Wiley & Sons.

Lieb, Elliott H. and Jakob Yngvason (1999). “The Physics and Mathematics of the Second Law of Thermodynamics”. In: *Physics Reports* 310, pp. 1–96.

Mandl, F (1988). *Statistical Physics.* 2nd ed. John Wiley & Sons.

Mrugała, R. (1978). “Geometrical Formulation of Equilibrium Phenomenological Thermodynamics”. In: *Reports on Mathematical Physics* 14.3, pp. 419–427.

— (2000). “Geometrical Methods in Thermodynamics”. In: *Thermodynamics of Energy Conversion and Transport.* Ed. by Stanislaw Sieniutycz and Alexis De Vos. New York, NY: Springer New York, pp. 257–285.

Muller, I. (1985). *Thermodynamics.* Pitman.

Myrvold, Wayne (2017). “Probabilities in Statistical Mechanics”. In: *The Oxford Handbook of Probability and Philosophy.* Ed. by Alan Hajek and Christopher Hitchcock. Oxford University Press.

Myrvold, Wayne C. (2011). “Statistical Mechanics and Thermodynamics: A Maxwellian view”. In: *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 42.4, pp. 237–243.

— (Aug. 2020). “The Science of $\Theta\Delta^{cs}$”. In: *Foundations of Physics* 50.10, pp. 1219–1251.

Norton, John D. (2016). “The Impossible Process: Thermodynamic Reversibility”. In: *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 55, pp. 43–61.

Partington, James Riddick (1913). *A Text-book of Thermodynamics.* Constable & Company Ltd.

Pierce, Felix J. (1968). *Microscopic Thermodynamics.* International Textbook Company.

Planck, Max (1903). *Treatise on Thermodynamics.* Trans. by Alexander Ogg. Longmans, Green & Co.

Pooley, Oliver (2021). "The Hole Argument". In: *The Routledge Companion to the Philosophy of Physics.* Ed. by E. Knox and A. Wilson. Routledge, pp. 145–159.

Psillos, S. (1999). *Scientific Realism: How Science Tracks Truth.* Philosophical issues in science. Routledge.

Ramsey, Norman F. (July 1956). "Thermodynamics and Statistical Mechanics at Negative Absolute Temperatures". In: *Phys. Rev.* 103 (1), pp. 20–28.

Riedi, P.C. (1988). *Thermal Physics: An Introduction to Thermodynamics, Statistical Mechanics, and Kinetic Theory.* Oxford science publications. Oxford University Press.

Roberts, Bryan W. (2020). "Regarding 'Leibniz Equivalence'". In: *Foundations of Physics* 50.4, pp. 250–269. DOI: `10.1007/s10701-020-00325-9`.

Roberts, J. K. and A. R. Miller (1960). *Heat and Thermodynamics.* Blackie & Son Ltd.

Saunders, Simon (2006). "On the explanation for quantum statistics". In: *Studies in the History and Philosophy of Modern Physics* 37, pp. 192–211.

— (2013). "Indistinguishability". In: *The Oxford Handbook of Philosophy of Physics.* Ed. by R. Batterman. Oxford University Press, pp. 340–380.

— (2018). "The Gibbs Paradox". In: *Entropy* 20.8.

Schrödinger, Erwin (1989). *Statistical Thermodynamics.* New York: Dover.

Sears, Francis W. and Gerhard L. Salinger (1975). *Thermodynamics, Kinetic Theory, and Statistical Thermodynamics.* Addison-Wesley.

Smith, R. A. (1952). *The Physical Principles of Thermodynamics.* Chapman & Hall Ltd.

Swendsen, R. H. (2018). “Probability, Entropy, and Gibbs’ Paradox(es)”. In: *Entropy* 20 (450).

ter Haar, D. (1954). *Elements of Statistical Mechanics.* Reinhart & Co.

— (1966). *Elements of Thermostatistics.* Holt, Reinhart & Winston.

Tisza, László (1961). “The Thermodynamics of Phase Equilibrium”. In: *Annals of Physics* 13.1, pp. 1–92.

Tribus, Myron (1961). *Information theory as the basis for thermostatics and thermodynamics.* D. van Nostrand Company.

Truesdell, C. (1980). *The Tragicomical History of Thermodynamics, 1822–1854.* Studies in the History of Mathematics and Physical Sciences. Springer New York. ISBN: 9780387904030.

Tschoegl, N.W. (2000). *Fundamentals of Equilibrium and Steady-State Thermodynamics.* Elsevier Science.

Uffink, Jos (2001). “Bluff Your Way in the Second Law of Thermodynamics”. In: *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 32.3, pp. 305–394.

— (2007). “Compendium of the Foundations of Classical Statistical Physics”. In: *Philosophy of Physics.* Ed. by Jeremy Butterfield and John Earman. Handbook of the Philosophy of Science. Amsterdam: North-Holland, pp. 923–1074.

van Kampen, N. G. (1984). “The Gibbs Paradox”. In: *Essays in theoretical physics in honour of Dirk ter Haar.* Ed. by W. E. Parry. Oxford: Pergamon Press, pp. 303–312.

van Lith, Janneke (2018). “The Gibbs Paradox: Lessons from Thermodynamics”. In: *Entropy* 20.5.

Weatherall, James Owen (2018). “Regarding the ‘Hole Argument’”. In: *British Journal for the Philosophy of Science* 69.2, pp. 329–350.

Wightman, Arthur S. (1979). *Introduction to Convexity in the Theory of Lattice Gases by Robert B. Israel*. Vol. 64. Princeton University Press.

Wills, James (2021). “Representing Mixtures”. manuscript.

Zemansky, Mark W. and Richard H. Dittman (1997). *Heat and Thermodynamics*. McGraw-Hill.