Neo-Nagelian Reduction

A Statement, Defence, and Application.

Foad Dizadji-Bahmani

A thesis submitted to the Department of Philosophy, Logic and Scientific Method of the London School of Economics and Political Science for the degree of Doctor of Philosophy, July 2011.

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. 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 the prior written consent of the author. I warrant that this authorization does not, to the best of my belief, infringe the rights of any third party.

Foad Dizadji-Bahmani

Abstract

The thesis proposes, defends, and applies a new model of inter-theoretic reduction, called "Neo-Nagelian" reduction. There are numerous accounts of inter-theoretic reduction in the philosophy of science literature but the most well-known and widely-discussed is the Nagelian one. In the thesis I identify various kinds of problems which the Nagelian model faces. Whilst some of these can be resolved, pressing ones remain.

In lieu of the Nagelian model, other models of inter-theoretic reduction have been proposed, chief amongst which are so-called "New Wave" models. I show these to be no more adequate than the original Nagelian model.

I propose a new model of inter-theoretic reduction, Neo-Nagelian reduction. This model is structurally similar to the Nagelian one, but differs in substantive ways. In particular I argue that it avoids the problems pertaining to both the Nagelian and New Wave models.

Multiple realizability looms large in discussions about reduction: it is claimed that multiply realizable properties frustrate the reduction of one theory to another in various ways. I consider these arguments and show that they do not undermine the Neo-Nagelian of reduction of one theory to another.

Finally, I apply the model to statistical mechanics. Statistical mechanics is taken to be a reductionist enterprise: one of the aims of statistical mechanics is to reduce thermodynamics. Without an adequate model of inter-theoretic reduction one cannot assess whether it succeeds; I use the Neo-Nagelian model to critically discuss whether it does. Specifically, I consider two very recent derivations of the Second Law of thermodynamics, one from Boltzmannian classical statistical mechanics and another from quantum statistical mechanics. I argue that they are partially successful, and that each makes for a promising line of future research.

Outline

Contents

Acknowledgements

I thank Roman Frigg and Miklós Rédei for their patient and supportive supervision. Miklós has been a source of inspiration and I learned much from him. I am particularly indebted to Roman. He has provided many insightful comments on my writings, explained important ideas to me, and guided me through this process. I am extremely grateful to them both.

I thank all the staff and students at the Department of Philosophy, Logic and Scientific Method at the London School of Economics and Political Science, for creating the excellent academic atmosphere here. In particular, amongst the faculty, I thank Jason McKenzie Alexander, Luc Bovens, Nancy Cartwright, David Makinson, Katie Steele, Max Steuer, Charlotte Werndl, and John Worrall. In particular, amongst the PhD students, I thank Bengt Autzen and Chris Thompson, my year-mates, Sheldon Steed, Mauro Rossi, Marilena Di Bucchianico, Alice Obrecht, Ittay Nissan, Dean Peters, Ben Ferguson, Susanne Burri, and Seamus 'LaTeX' Bradley.

I received financial support from the LSE Research Studentship Scheme and a desk from the CPNSS. The latter allowed me to benefit from stimulating conversations with Isabel Guerra, Iñaki San Pedro, and Wolfgang Pietsch, amongst many others. Most importantly it allowed me to share many a great day with Conrad Heilmann.

I thank James Ladyman, Stuart Presnell, and Ioannis Votsis for encouraging me to pursue a PhD, and Jeremy Howick and Matt W. Parker for seeing me through. Special thanks go to Phillip Thonemann for introducing me to philosophy.

I thank all my friends and family for putting up with me whilst I wrote this thesis. Special thanks to my Berlin family, Ayshea, Bex, Lucy, Susannah, Ula, and Mimi. I would not have completed this thesis without Emma, Henry, and Shiv.

I am sorry that my father is not here to see me get this far; from what I understand he would have been proud. I thank him, in so far as that makes sense, for being a source of inspiration and strength to my mother, even after all these years.

Above all I thank my mother, Farideh Dizadji. She has sacrificed so very much for me, and, despite incredible adversity, has provided me with all the intellectual and emotional support I needed. I owe her everything.

For Farideh Dizadji.

Thesis Overview

In this thesis, I propose, defend, and apply a new model of intertheoretic reduction. I call it the *Neo-Nagelian* model of intertheoretic reduction.

Let us idealise somewhat and suppose that any model of intertheoretic reduction consists of a set of necessary and sufficient criteria for one theory to reduced to another. I make the distinction between two kinds of problems pertaining to any such model: *internal* and *external*. By *internal* problems I mean those to do with how clear and precise the criteria are, whether the criteria are consistent with each other, and whether they afford the putative aims of reduction. Contrast this with the *external* problem. The *external* problem is the problem of establishing what the aims of reduction are, and whether the criteria themselves are the 'right' ones. That is, whether in meeting the criteria, one theory *does* reduce to another. The failure to recognise this distinction has been detrimental to debates about reduction, or so I contend.

To solve the *external* problem, I propose the following: abstract a general model of reduction from a rational reconstruction of the derivation of the Boyle-Charles of thermodynamics from statistical mechanics. Motivating this proposal is the first task of chapter 1. I call the model of reduction that this yields the *Neo-Nagelian* model of intertheoretic reduction. As the name suggests, it is structurally similar to Nagel's well-known model of reduction. However, it differs in substantive ways.

Nagel's model is widely regarded as untenable, for, in terms of the aforementioned distinction, it suffers from various *internal* problems. Given the similarity of the two, it must be shown that the problems with Nagel's model do not adversely affect the Neo-Nagelian one. Thus, before setting out the Neo-Nagelian model, I first present Nagel's model of reduction and indicate exactly what its *internal* problems are.

I then proceed to the central part of chapter 1, namely detailing and defending

the Neo-Nagelian model of reduction. I start by giving a sketch of the derivation of the Boyle-Charles of thermodynamics from statistical mechanics from which I abstract the general model. What is the general model? Reduction, on this account, is an exercise in explanation: a Neo-Nagelian reduction affords an explanation of the empirical success of the reduced theory by the reducing theory. Moreover, I shall show how Neo-Nagelian reduction can also afford ontological simplification. In so doing, I show that the model avoids the aforementioned *internal* problems besetting Nagel's original model.

Is Neo-Nagelian reduction the best alternative to Nagelian reduction? In one sense it is the only tenable alternative for it is unique in solving the *external* problem. Bracketing this point, in chapter 2, I consider, so-called, 'New Wave' models of intertheoretic reduction. Those advocating such models contend that even the best version of Nagelian reduction is untenable; New Wave reduction is proffered as the right alternative. I show that none of the various New Wave models avoid the problems pertaining to Nagel's account. In fact, I show that they are decidedly worse in important respects. Thus, they are not a credible alternative to the Neo-Nagelian model.

It is often argued that 'multiply realized' properties threaten intertheoretic reduction. In chapter 3, I first examine Fodor's famous argument for this conclusion. I argue that it has no purchase against Neo-Nagelian reduction. I then broaden the discussion and consider whether multiply realizability undermines the explanatory import of Neo-Nagelian reduction or the ontological simplification that a successful Neo-Nagelian reduction affords. I argue that it does neither.

In chapter 4 I look more closely at the derivation of the Boyle-Charles law from statistical mechanics, and substantiate the claims I made about this derivation in chapter 1. One of the distinctive features of the Neo-Nagelian model I advocate is its stance on bridge-laws. The dominant view in the literature on reduction is that bridge-laws are, at least in the ideal case, property identities. I claim that this view is misguided, and I do by providing a detailed examination of a test case: the 'temperature - mean kinetic energy' bridge-law.

With the Neo-Nagelian model substantiated and defended, I then apply it to an important case: the putative reduction[1] of the Second Law of thermodynamics to Boltzmannian classical statistical mechanics. The Neo-Nagelian model provides a clear normative framework for considering this case and I conclude

[1]Worry not: the sense in which one can reduce laws as opposed to whole theories is detailed below.

that it is at least partially successful.

In chapter 5, I apply the Neo-Nagelian model to quantum statistical mechanics. There have been several recent attempts to reduce the Second Law of thermodynamics to quantum statistical mechanics. I show that the most promising amongst them is also partially successful.

Finally in chapter 5.6 I offer my conclusions and point towards areas for promising future research.

Chapter 1

Neo-Nagelian Reduction

1.1 Chapter 1 Introduction

Intertheoretic reduction is a perennial theme in philosophy of science; a topic that has been present since the very beginning of analytic philosophy of science. There is a striking variety of reductive claims. Some claim that the very *modus operandum* of science is reductive, others that the history of science is replete with reductions, others still that the putative exemplars of reduction in science are not reductions after all, and yet others that intertheoretic reduction is not possible. Tied up with intertheoretic reduction are the notions of ontological reduction and reductionism, where, roughly speaking, the former is the reduction of objects or properties to others, and the latter is the claim that all of science reduces to physics. Yet, before one can consider whether or not reductions are ubiquitous, numerous, few, or impossible; whether science aims at reduction; whether all of science does reduce to physics; and so forth, *one must first settle what it is for one theory to reduce to another.*

So what is it for one theory to reduce to another? Various models of reduction have been put forward. Prominent amongst them are: Kemeny and Oppenheim (1956); Schaffner (1967); Nickles (1973); Suppe (1974); Hooker (1981); Churchland (1985); Batterman (1995); Bickle (1998); Kim (2000). The most widely discussed model of reduction is due to Nagel (1961). Each of these models has been criticised as problematic in various ways.

It is important to distinguish between two kinds of problem. The first kind is what I will call *internal*, the second *external*.[1] Let us idealise somewhat and

[1]This is not substantively related to Carnap's (Carnap (1950)) distinction between internal

suppose that each model purportedly provides a set of necessary and sufficient criteria for one theory to reduced to another. By *internal* problems I mean those to do with how clear and precise the criteria are, whether the criteria are consistent with each other, whether they afford the putative aims of reduction, etc. Contrast this with the *external* problem. The *external* problem is the problem of establishing what the aims of reduction are, and whether the criteria themselves are the 'right' ones. That is, whether in meeting the criteria, one theory *does* reduce to another. For example, why think that the derivation of the laws of a theory from another constitutes reduction, or that reduction requires explanation, or ontological simplification? Unfortunately a distinction between *internal* and *external* problems is not made in the literature, and that has been, and continues to be, to the detriment of the clarity of debates, as shall become clear. A priority suggests itself here: first one ought to settle the external problem - i.e. one should determine what the 'right' criteria for reduction are - and then deal with the internal ones.[2] Yet, there is also interplay between the two: if a model of reduction fails *internally* then the question of its external validity, if you will, is seemingly irrelevant. But there may be various internally unproblematic models, which then prompts the question of which one is the correct one. Moreover, the *internal* problems with any given model may be surmountable, once the *external* problem has been settled.

1.2 The External Problem and Methodology

1.2.1 The External Problem

What is the solution to the *external* problem? That is, what is the correct set of criteria for one theory to reduce to another? This question gives way to another; what is puzzling here is that it is not even clear how this question is to be settled. What is the right method by which one ascertains the correct notion of reduction - what does it take to show that a particular notion of reduction is the correct or incorrect one? This a particular instance of a general philosophical problem: the problem of concept formation. How does one form concepts and justify that formulation, and in particular, how does one do that for concepts in philosophy

and external questions vis-á-vis linguistic frameworks, although readers familiar with Carnap's distinction might see an analogy.

[2]This presumes that there is a single 'right' model of reduction. More about this presumption shortly.

of science?[3]

To illustrate the problem of concept formation in this context consider the following. Suppose one were presented with a model of intertheoretic *reduction* in the form of a set of necessary and sufficient conditions. That is, a set of conditions which, if a theory satisfies them, (putatively) entails that the theory reduces to some other theory. How does one settle whether this is the *right* set of conditions? Someone arguing contra this definition would (try to) find a counterexample: either an actual case of a reduction which does not satisfy the formal conditions, thereby showing the conditions not to be necessary, or a case which does satisfy the criteria but is not an actual case of reduction, thereby showing that the conditions are not sufficient. The *grammar* of this is right.[4] However, it defers the substantial part of the problem, for one needs a criterion by which the (putative) counterexample is deemed an 'actual' case of reduction or not.

As I said, this problem is not particular to philosophy of science – it looms large in all of philosophy. The standard way out of this impasse is by recourse to intuition: the putative counterexample needs to be *intuitively correct*. For example, in ethics, moral concepts are rejected or refined in light of further intuitive examples, a process aimed at reaching 'reflective equilibrium'. It is, arguably at least, fitting that intuition plays a role in the forming of moral concepts.[5] Let us suppose that this both a compelling and tenable method when it comes to moral concepts. Can it be equally well applied in philosophy of science?

There are two points to consider. First, whether one has intuitions about philosophy of science concepts at all. Second, whether, given a positive answer to the first, any 'weight' ought to be given to them.

Can we be said to have intuitions about intertheoretic reduction, say? It seems to me that if one has intuitions about such a recondite term at all, one has them in virtue of either forays into philosophy of science in the first place.

[3]Notice that this problem does not just apply to *reduction* but other philosophy-of-science concepts. For example, *emergence*, *unification*, or *incommensurability*.

[4]And as the example below shows this is not an idle possibility, this is indeed what goes on.

[5]This is far too much of a caricature of conceptual analysis in ethical discourse, of course, but like all caricatures it also bears some resemblance to it too: reflective equilibrium is not the sole method for 'ascertaining' ethical concepts but it is an important one. The important point for the present concerns, is that that people have moral intuitions and that it is plausible, it seems to me, that these be used to form our moral concepts. A full discussion of this goes far beyond the present work. And whatever one might think about the position alluded to here, it is important to note that this problem receives a lot more attention in ethics, or more properly in metaethics, than the analogous problem does in contemporary philosophy of science.

Or one has them in virtue of engagement with specific scientific examples, which scientists themselves regards as reductions. If this point is well taken, then one may question just how much weight should be given to such intuitions. It might well be argued that this is the case with moral concepts too but the distinction is, I take it, that there are reasons to *want* intuition to play a constitutive role in the forming of moral concepts. Moreover, quite generally it seems to be a fact that moral discourse is a part of human activity outside the confines of academia, as opposed to philosophy of science or science itself.[6]

I think that this is an under-appreciated methodological problem in philosophy of science. To show that it is a genuine problem, consider the following. In a chapter entitled 'Reduction of Thermodynamics', Sklar (1993) considers the Kemeny-Oppenheim model of reduction (Kemeny and Oppenheim (1956)). His critique of it perfectly exemplifies this methodological problem. In order to see this, it is necessary to outline the Kemeny-Oppenheim model itself.

1.2.1.1 The Kemeny-Oppenheim Model of Reduction

For Kemeny and Oppenheim reduction is a particular kind of progress in science. They identify two kinds of scientific progress: an increase in factual information and improvement to theories which are "designed to explain the known facts and to predict the outcome of future observations" (Kemeny and Oppenheim, 1956, 7), and identify reduction as a sub-kind of the latter. Specifically, reduction is the "replacement of an accepted theory... by a new theory... which is in some sense superior to it." (ibid)

In the first instance the superior theory "should fulfil the role of the old one, i.e., that it can explain (or predict) all those facts that the old theory could handle." (Kemeny and Oppenheim, 1956, 7) [7] However, they continue: "[W]e do not recognize the replacement of one theory by another as progress unless the new theory compares favourably with the old one in a feature that we can *very roughly* describe as its simplicity." (*ibid* org. emph.) Thus, by their lights the reducing theory is superior to the theory to be reduced, just in case the reducing theory strikes the best balance between strength and simplicity. Kemeny and

[6] All of this is intended to be non-value-laden.

[7] For Kemeny and Oppenheim, explanation and prediction are only pragmatically distinct: "[F]rom a logical point of view there is no difference between explanation and prediction. The distinction is a pragmatic one, depending on whether the fact deduced is already known or not yet observed." (Kemeny and Oppenheim, 1956, 8)

Oppenheim denote this notion of a 'best balance' by 'systematization'. Here is a quote summarizing their position:

> "As a first approximation we might say that the reducing theory should be simpler than the theory reduced. But this is not the complete answer. If the reducing theory is much stronger, it would seem reasonable to allow it some additional complexity. What our intuition tells us is that we must be satisfied that any loss in simplicity is compensated for by a sufficient gain in the strength of the body of theories. We need some measure that combines strength and simplicity, in which additional complexity is balanced by additional strength. Let us express this combined concept by talking about how well a theory is systematized... We will then require that the reducing theory be at least as well systematized as the theory reduced." (Kemeny and Oppenheim, 1956, 8)

What Kemeny and Oppenheim produce is a model which formalises the ideas in the quote above. The details of the formal model are not important for present concerns. But just by way of illustration, notice that there are, at least arguably, *internal* problems Kemeny and Oppenheim model. For example, it rests on a sharp distinction between observational and theoretical statements, a distinction which has fallen from grace in the current philosophical milieu.

1.2.1.2 Sklar's Criticism

In his chapter titled 'Thermodynamics and Reduction' Sklar considers the Kemeny and Oppenheim model of reduction and comes to reject it. On what grounds does he do so? Sklar's criticism is that the Kemeny and Oppenheim model fails to account for an important aspect of reduction. He writes:

> "If we look at *actual cases of reduction* in science, we find that there is always a close relationship between reduced and reducing theory at the level of theoretical structure." Sklar (1993, 335, emph. added.)

The argumentative structure of Sklar criticism is clear: there is a *necessary* feature of reduction - viz. theoretical structural similarity - which the Kemeny and Oppenheim model fails to account for, rendering it untenable. Clearly, Sklar's argument falls into the methodological problem indicated above. One *can* use

theoretical structural similarity as a (one) criterion for reduction but one would want to know why this is the correct criterion. Obviously it would beg the question to say that *actual* cases of reduction must have this feature.

A more charitable interpretation of Sklar's position is to take him to concur with Kemeny and Oppenheim as to what are the actual cases of reduction but disagree in their modeling of reduction based on them, i.e. we can suppose that they agree about the set of intertheoretic reductions but disagree about the characteristic features of this set. Indeed, Kemeny and Oppenheim start their paper in this vein:

> "There are many examples of reductions that have been achieved. For example, a great part of classical chemistry has been reduced to atomic physics; and the classical theory of heat has been reduced to statistical mechanics... *The difficulty lies in finding the essential features that such historical examples have in common.*" (Kemeny and Oppenheim, 1956, 7, emph. added)

So Sklar's criticism could be that whilst these are cases of reduction, the putative essential features as per Kemeny and Oppenheim's model are incorrect (or at least that their model is incomplete), for their model misses theoretical structural similarity. But even this more charitable interpretation does not allay the worry. Why, one might ask, is theoretical structural similarity an essential feature? Again, one has to fall back on intuition. Indeed, it is clear that intuition features twice: not only must it be intuitively-the-case that theoretical structural similarity is an essential feature of the relation that holds between pairs of reductive theories, but the very identification of these pairs as cases of reduction rests on intuition too.

It is a commonplace that all philosophical arguments eventually come down to intuitions; that shared premises are needed to make any philosophical progress. None of the above is intended as a rejection of recourse to intuitions *per se*. Again my point is only that in the *particular* case of intertheoretic reduction (and perhaps other 'specialist' terms in philosophy of science) it is seems to me that little weight should be given to these intuitions, if one has them at all.

1.2.2 A Better Methodology

How can one solve the methodological problem illustrated in the previous section? Recall that the *external* problem is settling the right criteria for intertheoretic reduction. What are the right criteria for reduction then? To avoid recourse to intuition, I advocate, what might be called, a 'stipulative' solution. My proposal is to abstract a model of reduction from a rational reconstruction of the derivation of the Boyle-Charles law of thermodynamics from the kinetic theory of gases. That is, I take the derivation of the Boyle-Charles law of thermodynamics from the kinetic theory of gases as the sole *exemplar* of reduction, and abstract a model of reduction based on a rational reconstruction of this case.[8] The model of reduction that method yields, I shall call the Neo-Nagelian model of reduction. (More about the model shortly.)

Of course, intertheoretic reduction is a relation between *theories* so, strictly speaking, the derivation of the Boyle-Charles law from the kinetic theory of gases cannot be an exemplar of it. So the proposal put more precisely: if every law in the theory to be reduced (from here on the 'to-be-reduced' theory) can be derived in this manner - i.e. the manner in which the Boyle-Charles law is derived from the kinetic theory of gases - then that theory will have been reduced to the reducing theory. I shall sometimes speak of *reducing* a *law* in this sense.

There is much precedence that the relation between thermodynamics and statistical mechanics is a reductive relation which is just to say that many philosophers consider thermodynamics to reduce to statistical mechanics.[9] By extension, at least implicitly and in the aforementioned sense, the Boyle-Charles law reduces too. Yet, there are those who deny that this is a genuine reduction.[10] As Bishop and Atmanspacher put it: "it is a fabled reductionist legend that thermodynamics can be reduced to lower-level physics descriptions." (Bishop and Atmanspacher, 2006, 1755) Of course, to claim (or deny!) that the derivation of the Boyle-Charles law *is* reductive would just be a lapse back into the methodological problem above. For, in either case, one would need to be clear about what notion of reduction is underpinning the claim. This requires a solution to the external problem, and it

[8] 'Classical statistical mechanics' is abbreviated 'CSM'. In chapter 5 I look at the case of quantum statistical mechanics, which is abbreviated as 'QSM'.

[9] c.f. Kemeny and Oppenheim (1956), Nagel (1949) and (1961), Schaffner (1967), Churchland (1985), Hooker (1981), Bickle (1998), Sober (1999), Shapiro (2000).

[10] c.f. Sklar (1993), Ager et al. (1974), Causey (1972), Kim (2000), Bishop and Atmanspacher (2006).

is this problem with which we are presently concerned.

The important distinction is that my proposal is a *stipulative* one: the rational reconstruction of the derivation of the Boyle-Charles law from the kinetic theory of gases is *stipulated* to be a reduction. The question of whether it *is* a reduction, were it asked, would betray a conceptual confusion, for this question can only be meaningfully asked once a notion of reduction is fixed. Now, once we *have* abstracted the Neo-Nagelian model of reduction in this way, we can then apply it to the original case to find, unsurprisingly, that the Boyle-Charles does reduce to the kinetic theory of gases but such a claim is obviously not a *substantive* one.

Clearly a lot needs to be said about this proposal but let me first allay some potential worries:

> Q1: Doesn't this trivialise the reduction of thermodynamics to statistical mechanics?
>
> Q2: Why settle on the aims of statistical mechanics with respect to thermodynamics, and why take the derivation of the Boyle-Charles law from the kinetic theory of gases as constitutive of reduction?

As regards Q1, the answer is that it *does not* trivialise the reduction of thermodynamics to statistical mechanics. The rational reconstruction of the derivation of the Boyle-Charles law from the kinetic theory of gases *constitutes* the model of reduction, as I said, however, it remains an open question whether the rest of the laws of thermodynamics can be derived in the requisite sense from statistical mechanics. Thus it remains an open question whether thermodynamics does reduce to statistical mechanics. Indeed, as we shall see in chapters 4 and 5, there are many interesting and substantive issues in this respect.

As regards Q2, from a purely logical point of view the selection is arbitrary. In a sense it has to be, if we are to avoid recourse to intuition. (In so far as the selection does meet one's intuitions, then so much the better.) Any other pair of theories, or specific laws within those theories, could be used to construct a model of reduction and, again from a purely logical point of view, nothing prevents multiple models. One could have reduction in the given sense, or Newtonian mechanics-special relativity sense, and so forth. Each would be a different 'yard-stick' by which to 'measure' the relation between other theories. There are non-logical considerations, however: the relation between thermodynamics

and statistical mechanics *qua* reduction has received a lot of philosophical attention. As the thorough-going discussion of it will show, there are many persistent misconceptions about it (*qua* relation) in the literature.

1.2.3 Prolegomenon

As I said, my proposal is to abstract a model of reduction from a rational reconstruction of the derivation of the Boyle-Charles law from the kinetic theory of gases. What this amounts to is best exemplified rather than characterised, but in broad terms the idea is as follows: take the *aims* and *methods* of the kinetic theory of gases apropos the Boyle-Charles law as stated, and practiced, by practitioners. I will then generalise this into a general model for reduction. Of course, practice cannot be taken uncritically at face value; what is called for is a rational reconstruction: the aims need to be made explicit and the methods shown to be sound with respect to these aims.

The model of reduction that this yields is, *prima facie*, similar to Nagel's model, which is why I call it 'Neo-Nagelian' reduction.[11] Moreover, I take over some of Nagel's nomenclature; coining new names for terms which play similar roles in each of the models strikes me as obtuse and needless. The effect of this, however, is to make the Neo-Nagelian model seem more similar to the Nagelian model than it actually is. The similarity between the two is basically structural and terminological - they differ substantively in various ways. Before delving into the details, it is worth giving a sketch of this.

First, Neo-Nagelian reduction solves the *external* problem as set out above.[12] Setting this issue aside, what are the similarities and differences between the two models? To reduce one theory to another, in the Nagelian sense, requires the derivation of the laws of the former from the latter, in conjunction with auxiliary assumptions and bridge-laws. Deriving the exact laws of the to-be-reduced

[11] That it is similar to Nagel's original model is not serendipitous. In a largely overlooked early paper of Nagel's - 'The Meaning of Reduction in the Natural Sciences' Nagel (1949) - he sets out to use the relation of thermodynamics and statistical mechanics as a 'basis for generalization' Nagel (1949, 111). (I suppose that the reason this paper has been overlooked is that it is from a rather obscure edited volume addition: *Science and Civilization*, Eds R.C. Stauffer, (1949).) It is evident that his seminal later works, where he outlines his general model of reduction, are based on the more careful considerations of this case. I should also add that my advocating the Neo-Nagelian model is not motivated by a want to rehabilitate the Nagelian model *per se* nor is the use of 'Neo-Nagelian' reduction, to be taken as such. I actually came across the aforementioned paper after deciding upon the given method. Also see Nagel (1979).

[12] More accurately, I solve the external problem via the method set out above which then yields the Neo-Nagelian model.

theory, however, has been deemed too stringent a requirement, and the so-called 'Schaffner-modified' Nagelian model has been advocated. On this view, it is sufficient for reduction to derive laws approximating those of the to-be-reduced theory, *ceteris paribus*. To reduce one theory to another, in the Neo-Nagelian sense, also requires the derivation of the laws of the to-be-reduced theory or laws approximating them, from the conjunction of the reducing theory, auxiliary assumptions and bridge-laws. It is in this sense that the two models are structurally and terminologically similar – the differences are in the details. On the Neo-Nagelian model there needs to be *warrant* - a notion which is explicated in detail below - for both the auxiliary assumptions and the bridge-laws. It is only by showing that there is *warrant* for the auxiliary assumptions and bridge-laws that a reduction of one theory to another can be explanatory, or so I shall argue.[13]

The Nagelian model (including the 'Schaffner-modified' version) suffers from various *internal* problems, in the above sense. Given the similarity of the two models, these problems potentially pertain to the Neo-Nagelian model too. However, I shall show that the Neo-Nagelian model actually avoids the *internal* problems from which the ('Schaffner-modified') Nagelian model suffers.

Finally, there is the issue of ontological simplification. Nagel himself was noncommittal about ontological simplification but his model has been appropriated by some to this end. I show that the means by which Nagelian reduction is taken to afford ontological simplification is misguided and untenable. I shall argue that a successful Neo-Nagelian reduction affords ontological simplification in a different way.[14]

I proceed as follows. In the next section, section 1.3, I set out Nagel's model of reduction and indicate the various *internal* problems with it. In section 1.4 I present the Neo-Nagelian model of reduction, and show that it avoids the problems besetting Nagel's model.

[13]This may not *seem* like much of a difference, but read on to see that it is!

[14]However, as I articulate below, ontological simplification is not a necessary condition for Neo-Nagelian reduction.

1.3 Nagelian Reduction

1.3.1 Nagel's Model of Reduction

Consider two theories, which I will refer to as the *to-be-reduced* and *reducing* theories.[15] Nagel's fundamental idea is that reduction consists in *deriving the laws* of the to-be-reduced theory from the laws of the reducing theory. This idea is formally captured by postulating two criteria for a successful reduction: 'Connectability' and 'Deriviability'.

'Deriviability' is the requirement that the laws of the to-be-reduced theory are to be derived from the laws of the reducing theory and some auxiliary assumptions. (These may be idealisations, 'limiting' assumptions, and the like - more about this shortly.) Schematically, we can think of the laws of the reducing theory and the auxiliary assumptions forming a set of premises from which the laws of the to-be-reduced theory are to be derived.

The laws of each theory are couched in terms of the theoretical predicates of the theory. Clearly, for it to be *possible* to derive the laws of the to-be-reduced theory from the reducing theory and auxiliary assumptions, it must be the case that the theoretical predicates of the former 'appear in' the set of premises. This is what the 'Connectability' criterion requires.

Nagel introduces two kinds of inter-theoretic reduction: 'homogeneous' and 'inhomogeneous'.[16] A homogeneous reduction is when the set of theoretical predicates of the to-be-reduced theory are a subset of the set of the theoretical predicates of the reducing theory. In this sense, 'homogeneous' reductions are ones where Connectability is straightforwardly satisfied. As an example of a homogeneous reduction, Nagel proposes that of Newtonian Mechanics to Special Relativity. Newton's laws are stated in terms of 'mass', 'force', 'acceleration', 'momentum' and so forth. Special relativity has these theoretical terms as a subset of 'its own' set of theoretical terms.[17]

Inhomogeneous reductions are those where the set of theoretical terms of the

[15]There are many different labels for these in the literature. Nagel used T_O and T_N respectively, where the indices 'O' and 'N' stand for 'old' and 'new'. But this is just a device which reflects that, by Nagel's light, many inter-theoretic reductions are between theories that have succeeded one another over time.

[16]Nagel sometimes refers to the latter as 'heterogeneous'. In the early 1949 paper, he also refers to this as a 'qualitatively discontinuous' reduction (Nagel 1949 107).

[17]Whether or not these are the *same* predicates is a controversial point. I am here just presenting Nagel's stance on the matter.

to-be-reduced theory, are *not* a subset of that of the reducing theory. To make the reduction possible – that is, to satisfy Connectability – so-called 'bridge-laws' are introduced. The *function* of bridge-laws is to connect the theoretical terms of the two theories. Formally, they indicate which terms of the reducing theory can be replaced by terms of the to-be-reduced theory in the derivation.

As an example of an inhomogeneous reduction, Nagel proposes that of thermodynamics (TD) to classical statistical mechanics (CSM). Indeed, for Nagel, this is an example of inhomogeneous reduction *par excellence*. The laws of thermodynamics are couched in terms of thermodynamic predicates such as 'temperature', 'entropy', 'heat capacity', and so forth. SM does not have these theoretical terms. Or at least, they are not obviously the *same* predicates. So, for example, 'entropy' does appear in SM but, by Nagel's lights, it is not the same as thermodynamic entropy. To be able to derive the laws of TD from SM, bridge-laws are needed. Nagel's exemplar is the bridge-law connecting temperature to mean kinetic energy.[18]

In summary, the laws of the to-be-reduced theory are to be derived from the conjunction of the laws of the reducing theory, various 'auxiliary assumptions' and, if needed, bridge-laws. If indeed they are, then the former theory is reduced to the latter. Or, put another way, one has a successful Nagelian reduction of the former to the latter if laws of the former can be derived from the latter, various auxiliary assumptions and bridge-laws.

For Nagel, reduction affords explanation. Indeed, the reduction of one theory to another *is* an explanation of the former by the latter, given the *deductive-nomological* model of explanation, which Nagel advocated. Nagel also considered whether reduction affords ontological simplification but was not committed to this.[19]

1.3.2 Nagelian Reduction: Internal Problems

In this section, I set aside the *external* problem and examine the various *internal* problems pertaining to Nagel's model. Some of these, I will argue, are not problems after all but some are, and are in fact pressing. With these in place, I shall go on to set out the Neo-Nagelian model of reduction, which, I shall argue, avoids the problems from which the Nagelian model suffers.

[18]This is the most widely discussed bridge-law in the philosophical literature on reduction. I will say more about it in the coming chapters.

[19]I consider the question of ontological simplification in section 1.4.7.

Nagelian reduction has, as I said, various *internal* problems. The problems can be categorised into three kinds: the first pertain to the general framework, the second to 'Connectability', and the third to 'Derivability'.[20] I shall show that, whilst some of them can be readily dealt with, others persist and importantly, they are not ameliorated by Schaffner's modification either. In short, even the Schaffner-modified Nagelian model is untenable.

1.3.2.1 Framework

There several alleged problems with the framework in which Nagel's model is couched. One set of these might be dubbed the 'syntactic view of theories' problem. Nagel was one of the leading logical empiricists and advocated what came to be called the 'syntactic view', or the 'received view', of theories.[21] Roughly speaking, on this view, a theory is taken to be an axiomatic system; theories express theorems stated in a formalized language. One of the important distinctions of this view is the sharp distinction between observational and theoretical statements. (For classic statements and defenses of this view of theories see Nagel (1961) and Carnap (1967).)

The 'received view' came to be widely rejected in the philosophy of science community. In part, this was due to the rejection of logical empiricism more widely. Critics of the syntactic view argued, amongst other things, that theories cannot be fully expressed in a formal language, that the requisite account of the relation between language and the World is not forthcoming, and that the distinction between observational and theoretical terms is not tenable. The so-called semantic view of theories[22] came to replace the syntactic view as the dominant position in philosophy of science. (cf., for example, Suppe (1989))

The problem for Nagelian reduction is then this: it is based on an untenable framework, and, as such, it is itself untenable. Statements to this effect can be found in Churchland (1985) and Bickle (1998). But there are, at least, three reasons which undercut this problem.

The first is that the debate between the syntactic and semantic 'schools' is certainly not settled and there are controversies on both sides.[23] This is in itself

[20]This is not an exhaustive list but it does contain the most prominent arguments.

[21]The 'received view' is more accurately a further specified position 'within' the syntactic camp but I use the term interchangeably here.

[22]In this context, 'semantic' is (to be) understood in the model-theoretic sense.

[23]For a recent discussion see Chakravartty (2001).

suggests that it would be premature to simply discard the Nagelian model in light of this.

Second, there is a danger of a kind of 'guilt by association'. Even if there are reasons to reject the syntactic view of theories, it is not obvious this necessitates a rejection of the Nagleiam model. Surely we would want to know what is supposedly wrong specifically with the model. In short, the inference from its being 'based on' an untenbable framework to itself being untenable is questionable. For example, were one to come to reject the syntactic view on the grounds that it the observational-theoretical distinction is untenable, this would not, I suggest, undermine the Nagelian model for nothing in the model rests on this distinction.

Most importantly, the Nagelian model is 'recovered' within the semantic view of theories anyway. The most thoroughly worked out program in the semantic 'school' is the so-called 'Munich School'. In their definitive statement of the position Balzer et al. (1987), Balzer, Moulines and Sneed provide a version of Nagel's model within the structuralist mould.[24] I think it unfruitful to quibble whether this 'really' is Nagel's model. What criterion of identity could one use? To be sure, it is not cast in the same terminology but, well, that is the point. I take it that Blazer et al. having taken themselves to be giving a model-theoretic rendition of Nagel's model is sufficient to show that Nagelian reduction does not stand or fall with the syntactic view.[25] Thus I set aside this kind of problem with Nagelian reduction, and, by extension, Neo-Nagelian reduction. Let us turn to the other problems.

Whilst the Connectability criterion is conceptually prior to the Derivability criterion, for the sake of expositional clarity, it is preferable to consider the Deriviability first. So, I now consider two problems with Nagel's Derivability criterion: the 'Falsity' problem and the 'Exact Derivability' problem.

1.3.2.2 The 'Falsity' Problem

The to-be-reduced theory, usually an older theory, is false, and the reducing theory is assumed to be true. But we cannot derive a false theory from a true one, so it seems that the Derivability requirement cannot be met. Call this the 'falsity' problem. This is one of the criticisms put forward by Feyerabend (1965), and is noted by Churchland (1985) and Bickle (1998).

[24] Rantala (1991) a provides a neat summary of the key claims in the book in his review of it.
[25] I discuss this in greater length in chapter 2 when considering New Wave reductionism.

The solution to this problem comes by noting that the auxiliary assumptions are, strictly speaking, false: it is not the case that one is deriving a false 'conclusion' from from true 'premises' - the false conclusion is (validly) deduced from a set of premises, some which are (strictly speaking) false. Of course, the auxiliary assumptions need not be false; the point is that they can be and usually are.[26] (The issue of the falsity of the auxiliary assumptions is taken up in far greater detail below. Here I am just pointing out that the purely formal problem of validly deriving something false from something true is avoided.)

1.3.2.3 The 'Exact Derivability' Problem

One serious objection to Nagel's model is that Derivability is not going to satisfied in all but the most rare cases. The point is usually stated thus: for the vast majority of interesting cases it is simply not possible to deduce the exact laws of the to-be-reduced theory. If reduction consists in the deduction of the laws of the to-be-reduced theory then Nagelian reduction is *de facto* unrealisable. This is a point that was made very early on against Nagel's model by Feyerabend (1965) and has been repeated by many others since. (cf. Schaffner (1967) for an overview.)

Of course, the Nagelian could bite the bullet here: if the exact laws of the to-be-reduced theory cannot be derived from the laws of the reducing theory and auxiliary assumptions, then the former theory simply does not reduce. There will be an immediate rejoinder: there are pairs of theories where the exact laws of the to-be reduced cannot be derived from the reducing theory but *these are actual cases of reduction.* Yet, this rejoinder falls foul of the methodological problem discussed above, section 1.2. In short, this problem, if intended against the *external* validity of the Nagelian model, is unpersuasive.

However, there is a way to recast the problem as an *internal* one. One of the aims of Nagelian reduction is explanation: explanation of the to-be-reduced theory by the reducing theory. Intuitively this is matter of degree - *there may be better or worse such explanations.* But this is incongruent with Nagelian reduction, for it is categorical. It seems that deriving the laws of the to-be-reduced theory *approximately* affords some sort of explanation but this cannot, it seems, be captured by the Nagelian model.[27]

[26] On a related note, it may well be objected that the reducing theory is not true. Quite, but then there is no falsity problem in the first place.

[27] This is not quite right, as we'll see when we discuss Neo-Nagelian reduction shortly, but

So the problem is that 'exact derivability' is too strong a criterion for reduction - and I emphasise that this should be understood in the *internal* sense - and requires weakening. As it happens, Nagel himself (Nagel 1974) suggested some such weakening, although he did not provide much by way of detail.[28] Schaffner (1967) proposed a modification of Nagel's original model along these lines and his was more detailed.[29] For Schaffner, if the exact law of the to-be-reduced theory cannot be derived, then what is needed is a derivation of a *strongly analogous* law. Clearly 'strongly analogous' needs to be carefully explicated - something which he did not do - but he did place at least one constraint on it: the analog law must be empirically more adequate than the law of the to-be-reduced theory.[30] Given both that Nagel himself suggested a weakening of 'exact derivability' and the prominence of Schaffner's modification to the same effect, I shall just subsume this into Nagel's model of reduction. I refer to this as the 'Schaffner-modified' Nagelian model of reduction.

What, one might wonder, motivates the constraint that the analog law be more empirically adequate than the law of the to-be-reduced theory? Schaffner does not answer this question; it is stipulated that it ought to be. One of the upshots of Neo-Nagelian reduction I advocate is that it answers this question.[31]

There is another problem related to the problem of 'exact derivability': In what sense it is not possible to deduce the laws of the to-be-reduced theory? We have already seen that the laws of the to-be-reduced theory are not derived from the reducing theory alone: they are derived from the reducing theory and the auxiliary assumptions. So the claim that it is not *possible* to deduce the laws of the to-be-reduced theory from the reducing theory needs qualification.

I take it that the claim that it is (at least in some cases) not possible to derive the exact laws of the to-be-reduced theory is not meant as some sort of elliptical expression for a general impossibility proof - that is just absurd. Presumably, it is the claim that is not possible to deduce the laws given certain - say 'permissible' - auxiliary assumptions.[32] Yet remarkably this is nowhere articulated in the

grant the appeal of this for the sake of argument.

[28]This is pointed out in Richardson (2008).

[29]Also, Schaffner takes a different stance on bridge-laws. I return to this in section 1.3.2.4.

[30]A similar weakening is proposed by the *New Wave Reductionists*. This is discussed in chapter 2.

[31]In so doing, I will also show that there has been a fundamental confusion about the derivation of the approximate laws of the to-be-reduced theories. cf. section 1.4.2.

[32]Actually this can be further refined. *Strong Derivability Problem*: Given the permissible auxiliary assumptions and the laws of the reducing theory, it is demonstrable that one cannot

literature on reduction; at best it is tacitly assumed. As this has gone unnoticed, the following questions have found no answer: what are the constraints on the auxiliary assumptions? What counts as 'permissible' here? The Neo-Nagelian model of reduction provides answers.

1.3.2.4 Problems with Bridge-Laws

The other pressing internal problem for Nagelian reduction is the problem of bridge-laws. As with the auxiliary assumptions, Nagel does not say all that much about bridge-laws. Whilst their *function* is clear, their use is unjustified and their status unclear.[33]

The first problem with bridge-law is that of establishing, for want of a better phrase, where they come from and what determines their form. That is, how the relevant bridge-laws are ascertained. Why is it, for example, that temperature is 'associated'[34] with mean kinetic energy? And why is it that the well-known temperature - mean kinetic energy bridge-law has the particular functional form that it has? And moreover what justifies this association? The problem of where bridge-laws come from and their justification are inextricably linked, and, as such, I refer to them singly as the "Where-From" problem. It is interesting that this problem is rarely, if ever, explicitly discussed. Philosophers take as given the various bridge-laws that populate discussions about reduction and never discuss where they come from. In almost all the literature on reduction, *bridge-laws seemingly drop out of the sky!* That is, not only did Nagel not address this, but even those who have come after him, so to speak, have not done so. For example,

deduce the requisite laws. *Weak Derivability Problem*: Given the permissible auxiliary assumptions and the laws of the reducing theory, the requisite laws have not been deduced. I shall assume the 'weak' version.

[33]A well-known argument against Nagelian reduction is Feyerabend's 'incommensurabilty' argument (Feyerabend, 1965). The crux of the argument is Feyerabend's claim that the meaning of theoretical terms is determined solely within the theory in which they are embedded. The consequence of this is that the 'Connectability' criterion is impossible to satisfy according to Feyerabend and recourse to bridge laws is nonsensical. (It is worth noting that Feyerabend also argued against 'Derivability'. As Nagel correctly points out, one cannot maintain both these arguments simultaneously: that 'Connectability' is met is a pre-requisite (conceptually prior) for a rejection of 'derivability'. Two theories cannot be logically inconsistent if they are incommensurable.) In this, I contend, Feyerabend was wrong but I do not wish to settle this matter. It will be clear that the Neo-Nagelian account of bridge-laws avoids this problem because bridge-laws are not semantic claims at all, as is presupposed by Feyerabend. Interestingly, Schaffner (1993, 411-477) considers bridge-laws to be co-extensions. Co-extension is a semantic claim and is vulnerable to Feyerabendian criticism.

[34]'Associated' is to be neutral place-holding term for whatever the status of the bridge-law is.

Schaffner (1967) and (1993) says nothing about where bridge-laws come from and why the properties stand in the particular functional relationship that they do.

Another problem is the *status* of bridge laws.[35] What *kind* of statements are bridge laws? To return to the previous example, how is 'associated' to be interpreted in the temperature-mean kinetic energy bridge-law? Refer to this as the "Status" problem. It is this problem that has received almost all the attention.

Nagel considers three possible interpretations of bridge-laws: meaning equivalence, conventional stipulations, or assertions about matters of fact. (Nagel, 1961, 354-355). The third interpretation breaks down further. A bridge-law could be: an identity (i.e. each predicate could refer to the same property), nomic correlation (i.e. the predicates could refer to nomically correlated properties), or, finally, a brute correlation (i.e. the predicates could refer to properties that are non-nomically correlated). Nagel himself was noncommittal between these different options.

The first two interpretations have been all but dismissed: bridge-laws are, surely, not semantic claims or matters of mere convention, it is claimed. The debate has centered around the third interpretation and in particular about whether bridge-laws can be interpreted as identities. One problem with taking bridge-laws to be matters of fact, in whichever guise, is that justifying it seems difficult. As Nagel points out (*op. cit.* 356) we cannot test bridge-laws independently. Various non-empirical arguments have been put forward for and against taking bridge-laws to be identities, however. The prominent argument *for* identities has been from consideration of parsimony; the prominent argument against from consideration of multiple realizability. The problem of the status of bridge-laws is an outstanding internal problem for the Nagelian model.[36]

As set out in section 1.2.3, Neo-Nagelian reduction is structurally similar to the Nagelian model. On both accounts, roughly speaking, reduction consists in the derivation of the laws of the to-be-reduced theory from the reducing theory, auxiliary assumptions and bridge-laws. Thus, *prima facie*, Neo-Nagelian reduction faces these very same *internal* problems. After presenting the Neo-Nagelian model in the following section, I shall show how to deal with these problems. I

[35]Sometimes this is referred to a the 'logical status' of bridge laws. cf. Churchland (1985).

[36]There is also an interesting point about bridge-laws *qua* identities which is never discussed: From a purely logical point of view, bridge-laws *qua* identities are superfluous for reduction. If the predicates connected by a bridge-law refer to the same property then the bridge law serves only as a kind of pedagogical aide mémoire. On the other hand, if there are two properties that are correlated (nomically or or otherwise), the bridge-law is not superfluous logically speaking.

offer an answer to the question of what constraints need to be placed on the auxiliary assumptions. Moreover, I solve the problems with bridge-laws. Specifically, I shall show that solving the "Where-From" problem solves the "Status" problem too.

1.4 Neo-Nagelian Reduction

Recall from section 1.2.2 the solution to the *external* problem that I advocated: the Neo-Nagelian model of reduction is an abstraction of the rational reconstruction of the derivation of the Boyle-Charles law of thermodynamics from the kinetic theory of gases.

The plan for this section is as follows. In section 1.4.1, I shall first provide a sketch of the derivation of the Boyle-Charles law from the kinetic theory of gases. With this in place, I shall then present the Neo-Nagelian model of reduction in section 1.4.2, indicating the sense in which it is indeed a rational reconstruction of this derivation. The presentation of the derivation of the Boyle-Charles law in section 1.4.1 is really only a sketch. To substantiate the Neo-Nagelian model requires a more detailed examination of the derivation. This is provided in chapter 4; I hope that the reader will take my presentation here in good faith.

In presenting the Neo-Nagelian model, I shall also detail how the model differs from the Nagelian account and indicate how it avoids the *internal* problems pertaining to the latter. As shall become clear, to make good on the Neo-Nagelian model what is needed is a rehabilitation of the deductive-nomological (DN) model of explanation. To do so I propose a general framework for explanation, which I call *pluro-particularism*. I set the stage for *pluro-particularism* in section 1.4.3. I then present it and use it to rehabilitate the DN model in section 1.4.4. I then explicate the notion of *warranted* auxiliary assumptions and bridge-laws, which form the cornerstone of the Neo-Nagelian model, in sections 1.4.5 and 1.4.6. Finally, in section 1.4.7, I show how a successful Neo-Nagelian reduction affords ontological simplification.

1.4.1 Derivation of the Boyle-Charles Law - A Sketch

In this section I sketch the derivation of the Boyle-Charles law from the kinetic theory of gases. Of course, this derivation does not take place in a vacuum. What motivates this derivation, from the point of view of physicists? There are vari-

ous 'schools' in statistical mechanics each using different theoretical machinery.[37] However, what they all have in common is the aim of deriving the laws of thermodynamics. To be sure, deriving the laws of thermodynamics is not the *only* aim - it aims to give empirically more adequate laws and laws of broader scope - but the derivation of the laws of thermodynamics is central.[38] The derivation of the laws is taken to be explanatory. That is, the practitioners themselves regard derivations of this kind as providing explanation.[39] Why is this explanatory? One cannot take practice uncritically at face value. Showing that these kind of derivations are explanatory is central to the rational reconstruction I proffer below.

Exemplary of this kind of explanatory derivation, again by practitioners' lights, is the derivation of the Boyle-Charles law from the kinetic theory of gases. Here is a non-technical sketch of how this derivation goes.[40]

Consider the Boyle-Charles of thermodynamics. It states a functional relationship between three macroscopic properties, pressure, volume and temperature for a gas:

$$PV = kT \tag{1.1}$$

where P is pressure, V is volume, T is temperature and k is a constant. How is this law derived? Starting with the kinetic theory of gases, and adding various assumptions one constructs a model of the, so-called, 'ideal gas'. That is, one derives the Boyle-Charles from the conjunction of the kinetic theory and some

[37] cf. Sklar (1993) and Frigg (2008), for example. A potential worry is that the kinetic theory of gases is not, properly speaking, part of statistical mechanics. However, to insist that the kinetic theory of gases is not a part of it is, at best contrived, as statistical mechanics does not have a canonical formulation. I treat the kinetic theory of gases as a part of statistical mechanics. This is argued for in more detail in chapter 4.

[38] Open virtually any text book on statistical mechanics, and the first portion of the book will set out how to derive the laws of thermodynamics. More specifically, usually, first, kinetic theory of gases is used to derive some of the thermodynamic laws, such as the Boyle-Charles law. Second, the 'rest' of the thermodynamics is derived within the Gibbsian formalism. I shall also discuss the Boltzmannian approach. Again see chapter 4 for details.

[39] Here is Tolman attesting to this; his is one of the most widely cited textbooks on the subject: "The explanation of the complete science of thermodynamics in terms of the more abstract science of statistical mechanics is one of the greatest achievements of physics." (Tolman, 1938, 9). Contra Tolman, we shall see that it is not the case that we have a *complete* explanation of the thermodynamics. However, Tolman expresses the view that statistical mechanics *aims* at explaining thermodynamics.

[40] This example will be familiar to most readers hopefully. Again I remind readers that a detailed account of this derivation is provided in chapter 4.

assumptions:

> *Kinetic Theory of Gases*: a gas is a collection of particles obeying Newton's laws of motion, the definitions of pressure and kinetic energy.
>
> *Auxiliary Assumptions*: Space is isotropic; particles are point-particles, between which there are no attractive forces, and are reflected elastically from the walls; a velocity distribution for the particles $f(\vec{v})$ given by the so-called Maxwell-Boltzmann distribution.

The above schema is not in fact complete. The following equation relating temperature and the mean kinetic energy of the particles is then stipulated.

> *Theoretical Stipulation*: $T = \frac{2n}{3k} \langle E_{kin} \rangle$, where n is the number of particles and $\langle E_{kin} \rangle$ is the mean kinetic energy of the particles.

Once this equation is added to the derivation, the desired result follows, namely:

> *Conclusion*: $PV = kT$

This completes the sketch of the derivation of the Boyle-Charles law from the kinetic theory of gases. The general features of this derivation exemplifies the method by which the laws of the thermodynamics are derived, and, as I said, this kind of derivation is considered to be explanatory by practitioners.

1.4.2 Overview of Neo-Nagelian Reduction

What are the features of the derivation in the previous section? The derivation of the Boyle-Charles law is a matter of theoretical construction. It is derived from the kinetic theory of gases, some auxiliary assumptions and a further theoretical stipulation equating temperature to the mean kinetic energy of the particles. This is abstracted to form the Neo-Nagelian schema. The kinetic theory of gases is the reducing theory and the Boyle-Charles law is one of the laws of the to-be-reduced theory. In order to derive the Boyle-Charles law, one makes recourse to auxiliary assumptions and, as we saw, a further theoretical stipulation. The latter is the bridge-law. Thus Neo-Nagelian schema consists in the derivation of the laws of the to-be-reduced theory from the reducing theory, auxiliary assumptions and bridge-laws.

What of the aim of deriving the Boyle-Charles law in this manner? As indicated, practitioners consider some such derivation to be explanatory. The aim of statistical mechanics apropos thermodynamics, viz. explanation, determines the aim of Neo-Nagelian reduction; Neo-Nagelian reduction aims at explanation.[41]owever, the success of this derivation with respect to this aim cannot be taken for granted. The burden on the rational reconstruction of this derivation is to show that it is indeed explanatory.

The task now is to show that the Neo-Nagelian schema, viz. derivation of the laws of the to-be-reduced theory from the reducing theory, auxiliary assumptions and bridge-laws, satisfies the aim of reduction, viz. explanation. A successful NN reduction may also afford ontological simplification. However, ontological simplification is not a necessary condition for successful reduction; rather it is an upshot of a successful reduction. I return to the issue of ontological simplification in section 1.4.7; first let us concentrate on reduction and explanation.

As it stands the Neo-Nagelian schema does not afford explanation. In order to do so the auxiliary assumptions and bridge-laws need to be *warranted*. I explicate exactly what that means below. But with this notion introduced, I can now state the Neo-Nagelian model in full. Consider again two theories, the to-be-reduced theory and the reducing theory.[42]

> Neo-Nagelian (NN) reduction is primarily an exercise in explanation: *a successful Neo-Nagelian reduction affords an explanation of the empirical success of the to-be-reduced theory.*
>
> A NN reduction consists in deriving the laws of the to-be-reduced theory (or laws approximating them) from the conjunction of the reducing theory, and auxiliary assumptions and bridge-laws for which there is *warrant*.[43]

I now turn to filling in the details of NN reduction. The aim of NN reduction is an explanation of the empirical success of the to-be-reduced theory. How does NN reduction afford this? NN reduction affords explanation via the *deductive-nomological* model of explanation: one derives the laws of the to-be-reduced

[41]H

[42]Recall that in calling it the 'to-be-reduced' theory I am being neutral with regards whether the theory does or does not reduce.

[43]Again it is worth emphasising that, whilst structurally and terminologically similar to the Schaffner-modified Nagelian model, the NN model is substantively different, as is shown below.

theory (or laws approximating them) from the reducing theory, auxiliary assumptions and bridge-laws. In this sense, the reducing theory, auxiliary assumptions and bridge-laws are the explanans. Deriving the laws of the to-be-reduced theory (or laws approximating them) counts as having explained the degree of empirical success of the to-be-reduced theory precisely because it is a theory's laws that encode its empirical content.

The *deductive-nomological* model is not uncontroversial. In section 1.4.4, I set out a general philosophical framework for explanation - what I call *pluro-particularism* - and use it to rehabilitate the DN model of explanation. [44]

On the Schaffner-modified Nagelian account, and indeed alternative accounts of reduction such as New Wave Reductionism[45], there are two determinants of the success of a reduction: how accurately the laws of the to-be-reduced theory are derived and how counterfactual the auxiliary assumptions are that are used in the derivation. Clearly, and as noted in section 1.3.2.3, without putting restrictions on the auxiliary assumptions, any set of laws can be derived. Thus, judging the explanatory import of a reduction - i.e. how successful it is - only by how accurately one derives the laws of the to-be-reduced theory rather obviously misses the point.

The standard line then is that the success of a reduction be determined by the counterfactualness of the auxiliary assumptions used to derive the laws. So, the more counterfactual the auxiliary assumptions used to derive the laws of the to-be-reduced theory the 'worse' the reduction, the less counterfactual the better. *However, this confuses how 'good' a putative reduction is with how empirically adequate the laws of the to-be-reduced theory are!* From the point of view of the reducing theory (which is, for the purposes of reduction, supposed to be true) the to-be-reduced theory is usually, strictly speaking, false. Nonetheless, the to-be-reduced theory is, to some extent, empirically adequate. It is precisley this that needs explanation: it is the empirical success of the to-be-reduced theory which one wants to explain and one does this by deriving its laws from the reducing theory, auxiliary assumptions and bridge-laws. The counterfactualness of the

[44]There is an important difference between my use of the *deductive-nomological* model of explanation and what might be called the 'orthodox' version of it. On the orthodox account, a necessary condition for explanation is that the explanans are true. This obviously won't do here: the auxiliary assumptions in reductions are usually false! (For example, atoms being point-like particles, etc.) Thus on the account I advocate, the explanans need not be true but need to be *warranted*. (More about what it means for the auxiliary assumptions and bridge-laws to be *warranted* shortly.)

[45]I critically examine New Wave Reductionism in chapter 2.

auxiliary assumptions encodes 'how false', if you will, the laws are from the point of view of the reducing theory. But this cannot, therefore, be a measure of the success of the reduction itself.

NN reduction avoids this confusion. Clearly, the laws of the to-be-reduced theory are the explananda but the success of a reduction is *not* determined by how accurately the laws of the to-be-reduced theory are derived, *simpliciter*. Rather, there is a trade-off between how accurately one derives the laws of the to-be-reduced theory and auxiliary assumptions and bridge-laws used to derive them. What determines the success of a reduction is the extent to which the auxiliary assumptions and bridge-law used to derive the laws of the to-be-reduced theory are explanation supporting. What one wants from a reduction is a derivation of the laws from explananda which are explanation supporting; deriving the laws of the to-be-reduced theory, however accurately, from auxiliary assumptions and bridge-laws which are not, does not afford an explanation of those laws. In this sense that the cornerstones of NN reduction are the auxiliary assumptions and bridge-laws and explicating just what it is for them to be explanation supporting is one of the main tasks at hand. The particular sense in which auxiliary assumptions and bridge-laws are (need to be) explanation supporting, I call '*warrant*'.

There are various kinds of auxiliary assumptions: *idealisations*, *limits*, *dynamical assumptions*, and *initial conditions*. For a reduction to be successful the auxiliary assumptions need to be explanation supporting, as I said. Whether auxiliary assumptions *are* explanation supporting is a matter of degree, context-specific, and non-formal; I shall call this the *warrant* for the auxiliary assumptions. So when I speak of their being strong *warrant*, weak *warrant*, etc, for auxiliary assumptions, this just to express the degree to which the auxiliary assumptions are explanation supporting. In turn, NN reduction is also a matter of degree. We are concerned with *how successful* a particular reduction is, rather than whether a theory does or does not reduce. The other cornerstone of NN reduction are bridge-laws. Bridge-laws on the NN account are a particular kind of theoretical stipulation, which I shall call *coherence constraints*. This is in stark contrast to the all the interpretations of bridge-laws in the literature on reduction. Like auxiliary assumptions, bridge-laws need to be explanation supporting too. Again I refer to this as *warrant* for the bridge-laws. What *warrant* amounts to in both cases is detailed in sections 1.4.5 and 1.4.6.

An important difference between Nagelian reduction and NN reduction is

that, unlike the Nagelian which makes recourse only to the *laws* of the reducing theory, NN reduction uses the entire theoretical machinery of the reducing theory.[46] In particular, to characterise the notion of *warrant* one needs to consider the 'metaphysical picture' of the reducing theory. By 'metaphysical picture' I mean a specification of what the world would be like if the reducing theory were literally true. Or in *possible world semantics*, the possible world at which the reducing theory is true. One explains the empirical success of the to-be-reduced theory on the supposition that the reducing theory is true.[47] Call this the 'background supposition'. The 'background supposition' plays two important roles in NN reduction. First, it grounds the sense of *warrant* for some of the auxiliary assumptions, namely *idealisations*, and *dynamical assumptions*. Second, it grounds the *warrant* for the bridge-laws. Again, precisely how it does so I set out in sections 1.4.5 and 1.4.6.

Finally it is important to note that on the Neo-Nagelian account, reducing one theory to another is a matter of theoretical construction: one constructs a model in which the laws of the to-be-reduced theory are derived. Call this *reductive construction*. Thus, whether one theory reduces to another is also a theoretical matter not an empirical one. Looking back at the derivation of the Boyle-Charles, this point is clear: it is a part of, albeit not particularly taxing, theoretical physics. Empirical considerations do not arise in the derivation.[48]

So to summarise: NN reduction aims at explaining the empirical success of the to-be-reduced theory, supposing the reducing theory to be true (the 'background supposition'). To do this, one provides a *reductive construction*: one derives the laws of the to-be-reduced theory, or laws approximating them, from the reducing theory, auxiliary assumptions and bridge-laws. The success of a reduction is determined by the extent to which the explanans are explanation supporting. That is, there needs to be *warrant* for the auxiliary assumptions and bridge-laws.[49]

[46] Another way to put this is to say that NN reduction can take the whole of the reducing theory as part of its explanada and not just its laws. To be sure, it is the laws of the to-be-reduced theory which are the explanadum of reduction but this in no way requires that the explanada be restricted to the laws of the reducing theory!

[47] Of course, the reducing theory will, in general, not be literally true so, *ultimately*, the explanation that reduction affords is only as good as the reducing theory is verisimiliar.

[48] By extension, reductionism - the thesis that all of science reduces to physics - is also a theoretical thesis and not an empirical one. This will be shown to be of significance when we come to the discussion of multiple realizability in chapter 3.

[49] Why recover the laws of the to-be-reduced theory, as opposed to merely the empirical phenomena that it was successful covering? Deriving the laws of the theory in the requisite

Finally, what about the issue of ontological simplification? There is a prominent thesis that reduction affords ontological simplification via bridge-laws *qua* identities.[50] Call this the *simplistic ontological simplification* thesis or the SOS thesis for short. As we saw in section 1.3.2.4, there are two problems pertaining to bridge-laws in the Nagelian model: 'Where-From' and 'Status'. As I mentioned there, I contend that solving the "Where-From" problem solves the "Status" problem too. Bridge-laws on the NN account are *coherence constraints*, a particular kind of theoretical stipulation as I said. The salient point with respect to ontological simplification is this: bridge-laws on the Neo-Nagelian account are neither identities, nomic correlations nor brute correlations - bridge-laws are not metaphysically substantive in any sense. Consequentially, I forgo the SOS thesis. But this, in turn, seems to undermine any prospects for ontological simplification. In section 1.4.7, I shall show how a successful NN reduction may afford ontological simplification in a different way. Adopting a Quineian meta-ontological position, I shall argue that a successful reduction shows that *if* the reducing theory is part of the best 'conceptual scheme' (in Quine's sense), then the to-be-reduced theory will not be. This affords ontological simplification because, at most, we will be committed to the ontology of the reducing theory to the exclusion of the to-be-reduced theory.[51]

Again, I emphasise that the above is just an overview of NN reduction. Clearly to give NN reduction any bite, and to show that this is not philosophical jargonisation!, I need to rehabilitate the DN model of explanation, and, in particular, explicate the notion of *warrant* for auxiliary assumptions and bridge-laws. To this I now turn.

1.4.3 Explanation

The notion of explanation is a much discussed and contested one in philosophy of science. The aim here is to first set out a broad philosophical stance on explanation - *pluro-particularism* - and use this to defend the *deductive nomological*

way constitutes explaining the (degree of) empirical success of the theory, precisely because it is a theory's laws which encode its empirical content. Recovering the empirical phenomena itself, so to speak, would be to explain a different explanandum; indeed different theories may account for the same empirical phenomena.

[50] Actually, as aforementioned, Nagel himself was noncommital as regards ontological simplification. Many advocates of his model have argued for this, however. Details in section 1.4.7.

[51] Again note: ontological simplification is not a necessary condition for successful reduction, it is a potential upshot of it.

explanation which underpins NN reduction.

In section 1.4.3.1 I set out the DN model of explanation. With this in place, I briefly consider other models of explanation in section 1.4.3.2 and in section 1.4.3.3 I consider some recent pluralist positions. I stress that considering these is not a needless digression: considering these other models of explanation is important to *pluro-particularism*. I will set out pluro-particularism in section 1.4.4. Finally I go on to examine how NN reduction affords explanation and in particular the notion of *warrant* in sections 1.4.5 and 1.4.6.

1.4.3.1 The Deductive Nomological Model of Explanation

There are many kinds of explanatory models in the philosophy of science literature. Prominent amongst them are the following: *Deductive-Nomological* (DN); *Statistical-Relevance* (SR); *Causal* (C); *Unificatory* (U). I shall briefly consider, and problems pertaining to, each in turn.

According to the DN model, the explanation of the explanadum - the proposition to be explained - consists in the derivation of explanadum from a set of propositions at least one of which is a law of nature.[52] The DN model allegedly faces problems several problems. Whilst it is to provide necessary and sufficient conditions for explanation, there are alledged counterexamples to it. Here I shall consider three prominent ones: the 'ink-pot' case, the 'birth-control pill' case, and the 'flagpole' case.

> The 'ink-pot' case: there is an ink-pot on a table. My knee happens to knock into the table causing the ink-pot to spill onto the floor. There is an explanation of this which is schematically similar to DN-type explanations but which does not involve a law of nature:
>
> I1: Knees bumping tables causes ink-pots to spill (given a set of further unspecified conditions);

[52]The emphasis on the explanada, and indeed the explanaans, being propositions as opposed to physical events (the latter being what are now generally accepted as appropriate explananda) is a birthmark of the DN model's heritage, viz. logical positivism. Still, it seems that to dismiss the model on such grounds would be a case of 'guilt by association': one can regard the explanation of the explanadum to consist in the derivation of a proposition which *refers* to the explanadum rather straightforwardly I suggest. (Likewise, the explanas and the required law.) The DN model has a probabilistic counterpart, the Inductive Statistical (IS) model, wherein the requisite law of nature is not a deterministic law but a probabilistic one. The criticisms of the DN model equally apply to the IS model.

I2: My knee did bump the table;

C1: So, the ink-pot spilt on the floor.

The point of the 'ink-pot' example is to show that the criteria for DN explanation are not necessary. In this example, the requirement that there be a law of nature in the explanation is unnecessary: purportedly this is a perfectly good explanation of why the ink-pot fell over in which we have not invoked a law of nature. (Of course, a lot turns on what a law of nature is, but the example is intended to have intuitive purchase.)

A second counter-example to the DN model is the 'birth-control' case.

The 'birth-control pill' case:

B1: All people who regularly take birth control pills do not become pregnant.

B2: John regularly takes birth-control pills.

C2: John does not become pregnant.

Here the DN conditions are met, in particular E1 is taken to be a law of nature. Yet, this fails to explain, intuitively speaking, why John does not became pregnant. He does not become pregnant because he is a man and not because he took the birth-control pills. Hence the criteria for DN explanation are not sufficient. Finally consider the 'flagpole' case:

The 'flagpole' case:

F1: Light propagates in a rectilinear fashion.

F2: The length of the shadow of a flagpole is x meters long and the sun is at an angle Θ to the horizon.

C3: The height of the flagpole is h meters.

Whilst explaining the length of the shadow of a flagpole from the law that light travels in a straight line, and the height of the flagpole and the Sun's position with respect to it, does seem explanatory, the above putative explanation does not. Again the criteria for DN explanation are shown to be insufficient for they do not capture the requisite asymmetry.

Other examples of the failure of the DN model abound. (cf. Salmon (1985) and Kitcher and Salmon (1989) for an overview.) It is concluded that that the DN model of explanation is untenable: it provides neither necessary nor sufficient criteria for explanation. Clearly, then, DN explanation needs rehabilitation if we are to use it to underpin the sense in which NN reduction affords explanation.

How then to proceed? I develop and defend a general framework for explanation - *pluro-particularism* - and I use this to rehabilitate the DN model. In order to do so, I first need to outline some other models of explanation. I do so very briefly in the next section, section 1.4.3.2

1.4.3.2 Other Models of Explanation

As I said, I need to outline some other models of explanation in order to develop and defend *pluro-particularism*. (Just why this is will become apparent.) This is the task of this section.

An alternative kind of explanation is the statistical-relevance model. In essence, the idea is that, rather than requiring a derivation of the explanadum, an explanation consists in specifying which factors are *statistically relevant* to bringing about the explanadum. This model seems to make sense of the 'birth-control pill' case: John's taking birth-control pills is not statistically relevant to his not becoming pregnant and hence an explanation is lacking. The SR model is also problematic however. It is said to be unintuitively uncompelling in at least two respects. (cf. Salmon (1985)). First, the *degree* of statistical relevance does not feature in the model. Second, and even more unintuitively, the same set of explanans can explain opposite explanada.

A third, but different problem, for the SR account is that it is a somewhat of a 'filler': in those cases where the DN model does not, intuitively speaking, 'fail', the DN model seems to be a better model of explanation. The SR model, instead, only fills the gaps when the DN model fails.

A third class of models of explanation are the causal mechanical (CM) model. Roughly characterised, an explanation consists in providing a causal-mechanistic account of the explanadum. As Woodward puts it:

> "We may think of the CM model as an attempt to capture the "something more" involved in causal and explanatory relationships over and above facts about statistical relevance." Woodward (2010, sec. 4)

To do justice to the causal-mechanistic model of explanation would require a serious digression. In particular, the central notions in this account, namely 'causal process' and 'causal interaction', are difficult to explicate briefly. Given the stance I advocate below, this would be unfruitful. But it is worth just quickly outlining the problems this view faces. The first is satisfactorily explicating just these notions, of course. Second, even if satisfactorily explicated, not all of what constitutes the causal process seems to explanatorily relevant. For example, a total proximate description by which a stationary billiard ball comes to move having been struck by another, involves, suppose, some blue chalk also flaring up.

> However, it is hard to see what in the CM model allows us to pick out the linear momentum of the balls, as opposed to these other features, as explanatorily relevant. (*ibid.*)

Third, the CM model seems to be committed to explanation at the lowest-causal level; that 'higher-level' putative explanations fail to be such because they cannot, in virtue of being 'higher-level', provide total proximate descriptions.

Unificatory models of explanation construe explanation as a unificatory enterprise. This is the model of explanation that was advocated by Friedman (1974) and has most notably been developed by Kitcher and Salmon (1989). To explain a certain phenomenon is to, roughly, show it to fall under a more general argumentative pattern.

> "Science advances our understanding of nature by showing us how to derive descriptions of many phenomena, using the same pattern of derivation again and again, and in demonstrating this, it teaches us how to reduce the number of facts we have to accept as ultimate." (Kitcher and Salmon, 1989, 523)

As with the other models of explanation, the unificatory model(s) is also thought to be problematic in various respects. Obvious questions arise about just what constitutes an argumentative pattern, and in what sense the explanadum 'falls under' the argumentative pattern. Moreover, even granting answers to these questions, there is further problem similar to the last problem mentioned for the CM model: the so-called "winner-takes-all" problem. (cf. Woodward (2010)) Again roughly, the problem is that, at least on Kitcher's unificatory model of

explanation, "only the most unified theory that is known is explanatory at all; everything else is non-explanatory." (Woodward, 2010, sec. 5) This, as Woodward says,

> "gives up on the apparently very natural idea... that an explanation can provide less unification than some alternative, and hence be less deep or less good, but still qualify as somewhat explanatory." (*ibid.*)

1.4.3.3 Pluralism

The survey above is certainly not comprehensive but it does indicate the four predominant positions pertaining to explanation. Interestingly, however, whilst the putative problems with each of these models are discussed at length in the literature on explanation, two, I contend, fundamental, points are hardly addressed at all: what is the method of adjudicating between different models of explanation? And why expect (or tacitly assume) a singular 'one-size fits all' model?

Just as with reduction itself, there is very little *discussion about* the methodology by which one 'arrives at' models of explanation. The tacitly adopted methodology is much like that of trying to reach 'reflective equilibrium' (cf. section 1.2.2): arguments *for* or *against* a certain model consist of presenting intuitively compelling examples and counter-examples.

Unlike the case of reduction, however, intuition finds a sound footing in the context of explanation. After all, unlike 'reduction', 'explanation' does feature in ordinary (i.e. non-philosophical) discourse.[53] Still, that there are such different models of explanation is, *ipso facto*, indicative of the disparateness of intuitions in this case.

This ushers in the second point, namely why it is tacitly assumed that there is just *one* correct model, in the first place. Why suppose explanatory monism? A recent statement of explanatory pluralism can be found in (Woodward, 2010). He notes that there is 'room' for a pluralism about explanation at least across different scientific disciplines:

> "Although the extreme position that explanation in biology or history has nothing interesting in common with explanation in physics seems unappealing (and in any case has attracted little support), it

[53] This observation suggests an erosion of a sharp distinction between 'ordinary' explanation and 'scientific' explanation.

> seems reasonable to expect that more effort will be devoted in the future to developing models of explanation that are more sensitive to disciplinary differences." (Woodward, 2010, sec. 6.3)

Whilst Woodward's suggestion is compelling in so far as it goes, viz. a pluralism about inter-disciplinary explanation, it does not go far enough, I suggest.[54] In fact, Woodword's position is arbitrarily stinted: why suppose intra-disciplinary explanation to be different? That is, what precludes a pluralism about explanation even *within* disciplines? In advocating inter-, but not intra-, disciplinary explanatory pluralism, Woodword seems to be committed to there being sharp boundaries between disciplines and uniformity within disciplines with respect to whatever methods are taking to accrue explanation in the first place. But this is surely wrong on both counts. *Looking* at the methods within disciplines and between them one *sees*, first, that different disciplines invoke the same kinds of explanations and, second, that within any particular discipline there are various kinds of explanations.

In contrast to the above I advocate both pluralism and particularism about explanation *tout court* i.e. I advocate pluralism and particularism about explanation inter- and intra- disciplines. For want of a better label, I call this *pluro-particularism*. The basic idea is that there are lots of kinds of explanation and that every putative instance of explanation needs to be assessed individually.

Before giving a positive characterisation of pluro-particularism, it is useful to give a negative one: to the best of my knowledge the kind of explanatory pluro-particularism I am advocating has not been advanced in the philosophy of science literature although there are those that advocate explanatory pluralism, from which I want to distinguish it.

Jackson and Petite, in their paper "In Defense of Explanatory Ecumenism" (Jackson and Pettit, 1992) argue against the position that, *ceteris paribus*, the more fine-grained an explanation, the better.[55] Their position is that for any given model of explanation, it is not always the case that the more fine-grained explanation is better than a coarse-grained one. On their view, one ought to be

[54] More recently, van Bouwel and Weber (2008) advocate explanatory pluralism and call it as much! Yet, their pluralism is of the Woodwardian kind: they argue for *inter*-disciplinary explanatory pluralism only.

[55] 'Ecumenism' does not obviously connote 'pluralism'. However, Jackson and Petite's use of 'ecumenism' (and indeed Førland's below) is not intended as monism but as a kind of pluralism in the given sense.

ecumenical with respect to fine-, and coarse-, grainedness, if you will. This is an 'explanatory pluralism' of sorts (although in any case, the labels do not matter much) but this is distinct from what I advocate.

Førland also picks up 'ecumenism' as a moniker in his (Førland (2004)) but defends a different position. Førland draws on the work of Peter Railton. He applies Railton's conception of explanation to the history. Railton's position is perhaps closest to what I have in mind and Førland provides a neat summary of Railton's position.

> "Railton's model of valid explanations is not restricted to causal or etiological explanations but admits several kinds of due-to relations. Railton argues that valid explanations need not be only causal but also can be structural or functional... One of the main features of Railton's account is that it accommodates many kinds of explanation that often have been regarded as incompatible, and combines them in one comprehensive model..." (Førland, 2004, 324)

Railton's model is closer to the position I wish to advocate at least in its pluralism. Yet the proffering of 'one comprehensive model', as Førland puts it, is unmotivated it seems to me: Railton whilst being in a certain sense pluralistic, is nonetheless trying to provide a comprehensive model. (Although I grant that this may just come down to a matter of presentation.) The significant difference between Railton's model and what I am advocating is this: it does not, crucially, emphasise particuralism.

1.4.4 Pluro-Particularism

Now I turn to the position I want to advocate: *pluro-particularism*. As I have indicated already, we are in methodologically murky waters when it comes to concepts in philosophy of science, for the very means by which particular concepts are defended or denounced is little discussed and understood. I have identified a tacit recourse to a kind of 'reflective equilibrium': using intuition to select exemplars of the relevant concept and abstracting a general characterisation of it from those exemplars. However, I have argued that this is an uncompelling methodology when it comes to the notion of reduction (at least), for there is little good reason to give 'weight' to intuitions in this case, if indeed one can be said to have intuitions about it at all. The situation is different in the case

of explanation because it features in ordinary discourse. We do have intuitions about what counts as explanation.[56]

The basic idea of pluro-particularism is that there a lots of kinds of explanation and each instance of explanation needs to be assessed individually. Notice that in the previous sentence I write 'and' and not 'but'. I do this because I do not think that the two conjuncts are in tension, although I write *this* because I can see that they might be *taken* to be! Indeed, the *interplay* between pluralism and particularism is crucial to my proposal as will become apparent below.

There are lots of kinds of explanation. By this I mean that there are different *explanatory schemata*. We have already encountered such explanatory schemata, namely those corresponding to the DN-model, SR-model, CR-model and unificatory models. I use 'schema' to indicate that these are general explanatory patterns for which there are various instances, i.e. that there are various explananda, some might be explained by the DN-model, some by the SR-model etc. However, none of these schemata guarantee explanation; none of these schemata provide necessary and sufficient conditions for explanation. Particularism requires that one look at each of the putative instances of explanation. For the various explanatory schemata, it is the *particulars* of those cases that need to be examined. Some explananda are explained in terms of the DN-model (i.e. for some explananda, there is a set of explanaans which jointly entail the explandum in a satisfactory way); some explananda are explained in terms of the SR-model (i.e. for some explananda, there is a set of explanaans which *are* statistically relevant to the explanandum); some explanada are explained by the CM-model (i.e. for some explanada, there is a causal mechanistic explanation)[57]; and so forth. In short there are intuitively compelling instances, sometimes also called 'paradigmatic' instances or 'exemplars', of each of these models. (In a certain sense, one wants to say that there *must* be intuitively compelling cases for the models would not have been suggested in the first place otherwise.) But we have also seen that they sometimes fail as explanation, as judged by intuition in particular cases. Here is the point about the inter-relation between pluralism and particularism: The value of having explanatory schemata does not lie in providing necessary and sufficient conditions for explanation: the schemata are more like a useful taxonomy,

[56]Put slightly more carefully: people have intuitions about (putative) instances of explanation. Given a putative explanation one has intuitions about whether it is a 'good explanation' or whether it '*really is* an explanation'. Unless otherwise indicated (either explicitly or by context) I shall mean a 'good' or 'real' explanation with 'explanation'.

[57]There are persuasive arguments for pluralism about causality itself. cf. Hitchcock (2003).

which allow for a better articulation of what is or is not explanatory, intuitively speaking, in any given case. *On this view, the explanatory schemata regiment the terms of the substantive debate but it is the particulars of the case that settle it.*

To give more of a feel for the particularism, reconsider the birth-control pill case. There is intuitive consensus that this not a case of explanation: John doesn't get pregnant because he is a man and the taking of the pills is neither here nor there! We can better articulate the intuition that there is a lack of explanation, however, by making recourse to explanatory schemata. The taking of the pills being 'neither here nor there' can be expressed in terms of statistical relevance: his taking of the pills is not statistically relevant to his not getting pregnant. One can tell a causal-mechanistic story (where 'story' is not intended pejoratively); indeed one can tell *several* such stories with varying degrees of detail about John's physiology. One can also 'locate' this instance within a broader class of phenomena: human males cannot get pregnant, or mammalian males cannot get pregnant and thus do justice to the unificatory intuition. The important point is that the terms of the debate are regimented by the explanatory schema but whether his taking birth control pills is explanatory depends on whether or not John's taking the pill *is* statistically relevant; whether there *is* a causal-mechanistic story; etc. The value of pluro-particularism, as I intend it, is to focus discussion about what is substantive in discussions of explanation away from a vain search for the general form of explanation.

I find it hard to see how there could be disagreement about the particulars of the birth control case but in other cases there may well be disagreements. For example, you and I might disagree about whether or not some factor is statistically relevant for some particular outcome. And notice that this disagreement may not be straightforwardly empirical in that the empirical specification of the problem may be what is at stake. A different example: Why does your cup of coffee cool? I might purport to explain it by citing the Second Law of thermodynamics. You might not be satisfied with this as a DN explanation, for you might not regard the Second Law of thermodynamics a genuine law of nature. Both these cases illustrate the kinds of disagreements that are interesting and substantive. Again, pluro-particularism, as I intend it, encourages the examination of these kinds of details instead of fruitless debates about the general form of an explanation.

1.4.4.1 Rehabilitating the DN Model

How pluro-particularism rehabilitates the DN model of explanation should now be obvious. *Sure*, the pluro-particularists says, *there are cases where the DN model goes awry* (e.g. the ink-pot case, the birth-control case) *but that does not mean that it is not an explanatory schema.* Specifically, the very idea that there are necessary and sufficient criteria for explanation *which the DN model fails to provide* is misguided. There are no such criteria; there are only explanatory schemata of which DN is one. It is the particulars of a putative DN explanation which need to be examined.

I have encountered two arguments against the pluro-particularist rehabilitation of DN explanation.

> (PPP1) *The Inconsistency Problem*: That some instances of the DN schema afford explanations but others do not is inconsistent: either the explanatory schema is a set of necessary and sufficient conditions for explanation or it is not!
>
> (PPP2) *The Too-Permissive Problem*: Pluro-particularism is *too permissive*: given this liberalism about explanation how can something fail to be an explanation? Pluro-particularism smacks of a kind of conceptual vagary which is best avoided!

PPP1 begs the question against pluro-particularism for, on that account, there is no set of necessary and sufficient conditions for explanation. One may be unpersuaded by the arguments for pluro-particularism given above - although I would implore you to re-examine your own monistic-bias! - but the 'Inconsistency' problem does not constitute a *distinct* problem for pluro-particularism. At most it is an expression of entrenchment to the effect that the arguments for it are uncompelling.

As regards PPP2, notice it is not the case that putative explanations cannot fail to be explanations on my account. Quite the opposite in fact: just because a putative case of reduction fits the DN schema does not entail that one has an explanation of the empirical success of the to-be-reduced theory. Rather the particulars of the case in hand need to be examined from close to. Pluro-particularism does not trivialise discussions about explanation. It is precisely aimed at moving the debate to the particulars, where the substantive issues are,

and away from idle disputes about the general form of an explanation, away from trying to establish *the* model of explanation.

Yet *these are remarks about* and *not arguments against* the claim that pluro-particularism is too permissive. I do not know how to argue against the claim because I do not know how it is argued for! That is, I do not know what the criteria for adjudication are here. If what I have said about pluro-particularism is well taken then one should be wary of the claim that there is a general criterion as to what counts as 'too permissive'. It is obvious to see why: some such criterion would indicate that there is a set of necessary and sufficient conditions for explanation after all.

Nonetheless PPP2 does prompt setting out some general desiderata for DN explanation. Of course, that these *are* desiderata for DN explanation rests on intuition, which just highlights that eventually all one can go on here is intuition. The point of setting out these desiderata is not to suggest otherwise; the desiderata articulate what those intuitions are.

D1: Explanation is not description.

D2: Explanation is asymmetric.

D3: Explanation is gradated not categorical.

D1. Explanation should be more than mere description - just describing the empirical success of a theory does not explain it. I concur with Glymour that explanation is in this sense "an exercise in the presentation of counterfactuals" (Glymour, 1970, 341) and not an exercise is the presentation of actual facts. As he concisely puts it:

> "One does not explain one theory from another by showing why the first is true; a theory is explained by showing under what conditions it *would* be true, and by contrasting those conditions with the conditions which actually obtain." (*ibid.*)

D2. Explanation should be asymmetric in the sense that for any putative explanation, the explanans ought to explain the explanandum but not vice-versa. It ought not to be possible to use the explanadum to explain the explanans. As a semantic claim this is obvious to the point of vacuity: that is what it *means* for something to be an explanan and an explanandum, respectively! The

worry is that formally one cannot distinguish between them: if DN explanation consists in derivation then it it seems that indeed one can use the hitherto assumed explanandum to explain one of the explanans. To do this one would simply 'reverse' the derivation: the explanan that one wishes to derive (and thereby explain) can be derived by what was originally the explanadum and the other explanans. (cf. the 'flagpole' case in section 1.4.3.)

D3. Explanation is not categorical but comes in degrees. Intuitively, there may be better and worse explanations for a certain explanadum, *even though each does explain it*. This is not to say that there are no explanatory assertions which are categorical. One does sometimes say that X explains Y, and the syntax of some such assertion suggests that explanation is categorical. But the grammar here is misleading: implicitly categorical sounding explanations are comparative (and explanations therefore are gradated).

The plan now is as follows. In the next section, section 1.4.4.2, I will show how the NN reduction as DN explanation meets the desiderata above. However, this will only be partially complete because, as will be shown, this turns on the *warrant* for auxiliary assumptions and bridge-laws. In section 1.4.5 I present how to think about *warrant* for auxiliary assumptions and in section 1.4.6 I do the same for bridge-laws.

1.4.4.2 DN Explanation and NN Reduction

At the start of this section we saw that statistical mechanics explains thermodynamics, at least by the lights of the practitioners. As set out above, my method is to start with the relation between thermodynamics and statistical mechanics, and to form the Neo-Nagelian model by a rational reconstruction of this case. But clearly taking practitioners claims about explanation at face value would be too easy. To substantiate the explanatory claim that NN reduction affords DN explanation, I show that it satisfies the desiderata D1 - D3.

D1. Explanation ought to be more than just description. NN reduction satisfies this desideratum as follows. The auxiliary assumptions are false and show under which conditions the laws of the to-be-reduced would be true. If all of the auxiliary assumptions *were* true, then the laws of the to-be-reduced *would* be true. Reconsider the derivation of the Boyle-Charles law: if gases *were* made of very many point-like particles of fixed mass with only kinetic energy governed by the laws of classical mechanics, then the Boyle-Charles law *would* be true.

Another way to think about this: the Boyle-Charles law is true at the possible world at which the kinetic theory of gases *and* the auxiliary assumptions are true.[58]

D2. Explanation should be asymmetric in the sense that for any putative explanation, the explanans ought to explain the explanadum but not vice-versa. We should not be able to 'reverse' the explanation. For example, we should not be able to explain the empirical success of the *kinetic theory of gases* from the Boyle-Charles law, and the same auxiliary assumptions and bridge-laws. Formally, this 'reverse' derivation goes through! NN reduction specifies what is wrong with this kind of 'reversed' derivation: there won't be *warrant* for the auxiliary assumption upon 'reversal'. Again, the derivation illustrates this: formally speaking if, as is the case, the Boyle-Charles law is derived from the laws of statistical mechanics and a set of auxiliary assumptions, then those very laws can be derived from the same set of the auxiliary assumptions and the Boyle-Charles. But the symmetry is broken by the *warrant*, or rather, the lack thereof, for the auxiliary assumptions used in the 'reversed' derivation. There is *warrant* for the auxiliary assumption when 'going from' the *kinetic theory of gases* to the Boyle-Charles law but not the other way round. In section 1.4.5 I shall argue for why this is so.

D3. Explanation is not categorical but comes in degrees. NN reduction satisfies this desideratum too. Just how good an explanation a reduction affords is determined by the best balance between the accuracy with which the laws of the to-be-reduced theory are derived and the degree of *warrant* for the auxiliary assumptions and bridge-laws. In the ideal case, one would derive the exact laws of the to-be-reduced theory from the reducing theory, and auxiliary assumptions and bridge-laws which are maximally *warranted*.[59] But the laws derived can be more or less accurate with respect to the laws of the to-be-reduced theory and there can be more or less *warrant* for the auxiliary assumptions. The measure for this is non-formal and context-specific but for all that not hopelessly vague nor lacking in conceptual cogency.

[58]"Isn't this world inconsistent? How can particles both have fixed finite volume and be point-like, for example?" We are to imagine the possible world at which each 'item' from the kinetic theory of gases which is in conflict with the auxiliary assumptions is, as it were, removed.

[59]Notice how this fits in with D2, the asymmetry desideratum: there is explanatory asymmetry precisely because there will in general be no *warrant* for the auxiliary assumptions needed, from the view point of (what would previously have been) the to-be-reduced theory. For details about the sense in which there is a tension here between the accuracy of the laws derived and the *warrant* for the auxiliary assumptions see the discussion in chapter 4.

1.4.5 *Warrant* For Auxiliary Assumptions

What counts as *warrant* for the auxiliary assumptions? First, this varies with the auxiliary assumptions: *warrant* differs for the auxiliary assumptions. Second, as a pluro-particularist (and here the emphasis is on the latter) I do not think that there is a formal and general measure for this. It is non-formal and context-specific; one has to look at the details of the case in hand. However we can exemplify the notion of *warrant* by drawing on the derivation of the Boyle-Charles law above.

There are four kinds of auxiliary assumptions:[60]

AA1: Idealisations

AA2: Limits

AA3: Dynamical Assumptions

AA4: Initial conditions

AA1: Idealisations are those assumptions which idealise the 'metaphysical picture' of the reducing theory. AA2: Limits are mathematical simplfications involved in the derivation. For example, one may drop certain terms in an equation as being 'negligible'. AA3: Dynamical assumptions are those about the dynamics of the system. Whilst we can think of Idealisations as idealising the 'metaphysical picture' of the reducing apropos its ontology, Dynamical assumptions idealise or make special probations for the reducing theory's dynamics. AA4: There are also assumptions about the initial conditions of the model - *reductive construction* - from which one wants to derive the laws of the to-be-reduced theory.

So what counts as *warrant* then for the different kinds of auxiliary assumptions? Let us start with warrant for the limiting assumptions, AA2. Were one to derive the same law from two sets of auxiliary assumptions, such that one derivation is more mathematically rigorous than the other, the former would constitute a better explanation. Batterman has written several influential papers on inter-theoretic reductions, (Batterman, 1995, 2000), involving asymptotic mathematical limits, i.e. non-converging limits. The thrust of Batterman's argument is that

[60]Notice that this list is not intended as mutually exclusive and exhaustive. The kinds of assumptions that a reduction involves may not fit neatly into the categories given here. However, I intend everything here to be suggestive enough to enable one to assimilate such assumptions into talk of *warrant*.

in these cases the putative reduction of the to-be-reduced theory is undermined. Whilst Batterman's categorical attitude towards reduction is misplaced at least by my lights, the intuition that there is something amiss in using such asymptotic limits is right: asymptotic limits are an example of an AA2 auxiliary assumption for which there is little *warrant*.

As regards initial conditions, the more general - the less 'special' - the initial conditions the better the explanation of the empirical success of the to-be-reduced theory based on it. Reconsider the derivation of the Boyle-Charles law but suppose that in order to have derived it it was insufficient just to assume the auxiliary assumptions we did but that we also had to assume that the gas particles start in restricted part of the container. This would undermine the explanatory import of the reduction, *ceteris paribus*, because it is in contradiction to the generality of the Boyle-Charles law. After all, the Boyle-Charles law is empirically adequate (to the degree that it is) for both gases that do and do not start in that special initial condition.

The *warrant* for Idealisations and Dynamical assumptions is more complicated to characterise. To do so it is helpful to again make use of possible world semantics. Call the possible world at which the reducing theory is true, PW_R.[61] There is *warrant* for these auxiliary assumptions (i.e. idealisations and dynamical assumptions) just in case, and to the extent to which, making them less counterfactual with respect to PW_R, would entail a derivation of laws which are more empirically adequate at the actual world, PW_A, than the ones that are derived.[62] To see why this is right, contrast it with what *warrant* is *not*: the *warrant* for these auxiliary assumptions is not determined by how counterfactual they are with respect to PW_R.[63] If this were the correct characterisation, then making the auxiliary assumptions less counterfactual with respect to PW_R, would mean that we would have more *warrant* for them and by extension we should have a better explanation of the laws of the to-be-reduced theory. But, of course, we would not have a better explanation of the laws of the to-be-reduced theory - indeed we would not have an explanation of them at all - for we would fail to derive the laws! The auxiliary assumptions *need to be* counterfactual because the laws of the to-be-reduced theory are, recall, strictly speaking, false from the point of

[61] That is to say, suppose the 'metaphysical picture' of the reducing theory.

[62] The counterfactualness of the auxiliary assumptions with respect to PW_R is itself something to be explicated in terms of possible world semantics. cf. Lewis (1969).

[63] This is the point that was flagged in section 1.3.2.3.

view of the reducing theory. The aim of the reduction was to explain the empirical success of a false theory, viz. the to-be-reduced theory, by deriving its laws, starting with the supposition that the reducing theory is true and in conjunction with the auxiliary assumptions and bridge-laws, So the counterfactualness of the auxiliary assumptions with respect to PW_R is not what grounds their *warrant*.

So what is the *warrant* for these auxiliary assumptions? They are *warranted* just in case *making them less counterfactual with respect to PW_R we obtain empirically more adequate laws*. In so doing we would show that we have not derived the laws of the to-be-reduced theory surreptitiously. Conversely, if increasing the veracity of the auxiliary assumptions with respect to PW_R entailed empirically *less* adequate laws, then they are not *warranted*.

This is best illustrated by an example. Consider again the derivation of the Boyle-Charles Law. One idealisation involved in the derivation is that the particles only have kinetic energy. Is this auxiliary assumption *warranted*? It is because were one to relax this assumption to include particles also having a weak interaction with one another - i.e. were one to make the assumption *less counterfactual* - one would derive a law relating the pressure and volume of a gas to its temperature which would be more empirically adequate. Indeed, what one would derive is Van der Waals equation, which is more empirically adequate than the Boyle-Charles law. Likewise the idealisation that the gas consists of mono-atomic particles. Were one to model the gas as di-atomic, one would arrive an empirically better law for the relation between the pressure, volume and temperature for an actual di-atomic gas.

Why is it that if in making an auxiliary assumption less counterfactual with respect to the reducing theory one gets empirically more adequate laws, then that auxiliary assumption is *warranted*? What is the idea behind this? If it is the case that in making an auxiliary assumption less counterfactual with respect to the reducing theory one gets empirically more adequate laws it shows us that the auxiliary assumption is on the correct 'explanatory trajectory'.[64]

Let me expand. By hypothesis, the reducing theory *itself* is at least approximately true. In particular, by hypothesis, it is more empirically adequate than the to-be-reduced theory. Thus, in terms of possible world semantics, that there is a straight line between the actual world, PW_A, to the possible world at which the to-be-reduced theory is true, PW_{TBR}, through the possible world at which

[64]Let me be clear that talk of possible worlds, explanatory trajectories and so forth is intended purely as a pedagogical ploy.

the reducing theory is true, PW_R. If the anteceedent of the conditional is true, then it shows us that the 'auxiliary assumption modified PW_R', all this PW_{R+AA}, lies on this trajectory. And this is exactly what we want.

Looking back at the examples in the previous paragraph, I think the intuitive appeal of the test is clear. For instance, that in making the auxiliary assumption about particles only having kinetic energy less counterfactual by modeling them as having potential energy too, and in so doing deriving more empirically adequate laws, we are assuring ourselves that the original assumption was not just ad hoc. It makes sense from the point of view of the reducing theory; it is not just surreptitiously used to yield the right result.

In summary: how good an explanation based on auxiliary assumptions is is determined by how *warranted* they are. But this cannot be measured by how counterfactual the auxiliary assumptions are, for they typically *need* to be counterfactual for the law that one is deriving is typically false. What one needs is an assurance that the auxiliary assumptions are not just ad hoc, that they are not put in surreptitiously merely to yield the right result. The test for this is that in making them less counterfactual one derives empirically better laws.

1.4.6 *Warrant* For Bridge-Laws

Let me first address the Where-From and Status problems. Bridge-laws play the same role in NN reduction as they do in Nagelian reduction: they 'connect' the predicates of the two theories. They appear as part of the explanans and their function in the derivation is to afford substitutions of the relevant predicates. Bridge-laws are a particular kind of theoretical stipulation, which I call a *coherence constraint*. They are not 'mere' or 'arbitrary' conventions, nor semantic claims, nor are the bridge-laws to be understood to be metaphysically substantive, which is to say that they are neither identities, nomic correlations nor non-nomic correlations. The Where-From problem is readily solved: the form of a bridge-law is determined by the particulars of the *reductive construction*, as we saw. So too is the Status problem: as I said, bridge-laws are particular kind of theoretical stipulation.

The substantive issue is this: Is it not too 'cheap' to stipulate the needed bridge-law? In what sense can such a stipulation be explanation supporting? That is, in what sense is there *warrant* for bridge-law *qua* theoretical stipulations? That is, in what sense are they explanation supporting?

The *warrant* for bridge-laws is determined by two things: *formal consistency* and *conceptual fit*. Put normatively, in order for a bridge-law to be *warranted* one must show that the properties it connects are formally consistent with one another and that their association fits conceptually, where this is determined with respect to the to-be-reduced and reducing theories respectively.

Reconsider our derivation and the bridge-law connecting temperature with mean kinetic energy. This bridge-law is *warranted* because there is both *formal consistency* and *conceptual fit* between temperature and mean kinetic energy. Roughly speaking, there is *formal consistency* because the properties associated with one another, viz. temperature and mean kinetic energy, via the bridge-law share the relevant formal properties. Put colloquially, they 'behave' in the same way. For example, they are both extensive properties, and both decrease as a function of pressure, and so forth. But *formal consistency* is not enough - one also needs to show that there is *conceptual fit*.[65] In this case there is for temperature is directly proportional to internal energy of an isolated system.

The moniker 'coherence constraint' should now have intuitive appeal: it connotes both that there is *formal consistency* and that there is *conceptual fit*. So the form of a bridge-law is determined by the particulars of the *reductive construction* and it is *warranted* just in case, and to the extent that there is *formal consistency* and *conceptual fit*.

The status of bridge-laws is that of theoretical stipulation, not an independently testable empirical hypothesis. It is often claimed against Nagel's original model that the use of bridge-laws undermines the explanatory import of a reduction. There are two claims to this effect. The first is that the very use of bridge-laws undermines explanation for where bridge-laws come from is entirely inexplicable - they are after all 'alien' to each theory. The second is that even if we somehow resolve the previous problem, anything short of bridge-laws qua identities is going to leave an explanatory gap: correlations or nomic connections between the properties of the two theories need an explanation themselves. Thus the general impression one gets in reading the literature on bridge-laws is that they either are invented by philosophers of Nagelian pursausion to suit their ends or alternatively that they must in some sense be factual claims about the relation between properties postulated by the theories - a fact that needs to be discovered - in which case they had better be identity claims to save our explanatory

[65] cf. the discussion of the so-called 'spurious reduction problem' in chapter 2.

blushes. But these claims are misplaced with respect to NN reduction, as the above shows. Bridge-laws are non-trivial theoretical stipulations which express a formal consistency given the particulars of the *reductive construction* and they do not undermine explanation precisely because there is *conceptual fit*.

(Even bracketing that, however, the claim that anything short of identities undermines explanation is very confused. First notice that the assertion that anything short of identities undermines explanation erroneously presupposes a categorical stance with respect to explanation in contradiction to desideratum 3. The obvious response is to say that a reduction involving bridge-laws qua identities is *better than* one where those bridge laws are correlations or nomic connections. Yet this is not true either: if the bridge-law *is* a factual claim, then only one of the possibilities can be true of course i.e. by hypothesis, the bridge law truly expresses either an identity, nomic connection or correlation. But which ever it is, it is as explanatory in each case. For example, if it *is true* that the properties are correlated then that is as explanatory as if they were identical or nomically connected, as each of these would be true to the exclusion of the others! It may be that on evidential grounds one cannot distinguish between correlations, nomic connections and identities and in this situation one may advocate bridge-laws qua identities on the ground of parsimony - better to have one property rather than two! - but this is irrelevant when it comes to explanation. That is, one may have motivations for advocating identities but considerations of explanation is not one. See also the discussion about multiple realizability in chapter 3.)

1.4.7 Ontological Simplification

The aim of NN reduction is explanation, and in the previous sections I have argued just how it affords this. However, I also indicated that NN reduction may afford ontological simplification. Before considering how so, let me recap briefly how ontological simplification is 'traditionally' thought to be had. On the Nagelian account, reduction affords ontological simplification via bridge-laws *qua* identities. This is what I called the *simplistic ontological simplification* thesis (SOS). As argued in the previous section (and as will bare out in more detail in chapter 4) bridge-laws are not metaphysically substantial in any sense - they are a particular kind of theoretical stipulation, viz. *coherence constraints*. Yet, as aforementioned, forgoing SOS seems to undermine any prospects for ontological simplification. In this section, I argue that a successful NN reduction may afford

ontological simplification in a different way.

Quine argued that we should be ontologically committed to whatever we quantify over in our "best conceptual scheme" (Quine, 1948). The Quinian dictum is well-known: "to be is to be a value of a bound variable" (Quine, 1948, 35) in our best conceptual scheme, where this is understood as just that scheme that does "justice to science in the broadest sense." (*ibid.*) There is much debate in meta-ontology and it is certainly not the case Quine's position is accepted by all. (cf. for example, Azzouni (1998) critical commentary.) I think that Quine's position is the right one but I shall not argue for it here - I shall just adopt it.[66] When the dust settles, whether or not NN reduction affords ontological simplification stands or falls with the Quinian meta-ontology, therefore.

Whilst Quine's position is, I contend, a compelling conceptual analysis of what it is for something to exist, it is not epistemologically very helpful. Quine says very little about what it is to do justice to science in the broadest sense and general criterion for this is not forthcoming. Whilst Quine left this largely to intuition, I suggest that there is an argument to be made for the following conclusion: a successful NN reduction of one theory another entails that the reduced theory is not part of our best conceptual scheme.

Suppose that we have a successful NN reduction of the to-be-reduced theory to another, the reducing theory. In this case, if the reducing theory is part of our best conceptual scheme, then the to-be-reduced theory is bound not to be, I suggest. Why so? For two inter-related reasons. We have an explanation of why the to-be-reduced theory is empirically successful (to the degree that it is) based on the supposition that the reducing theory is true.[67] This is important in this context because it allows us to use the to-be-reduced theory *without* being ontologically committed to it. Furthermore, the reducing theory is more empirically adequate - *ex hypothesi*, making the relevant auxiliary assumptions involved in the *reductive construction* less counterfactual entails more empirically adequate laws. Put in more colloquial terms: if we have a successful NN reduction of a theory, then we have (by definition) a successful explanation of why that theory is as empirically successful as it is, despite being false from the point of view of the reducing theory. This allows us to use the theory without being ontologically committed to it and

[66] I hope to defend Quine's position in future work. In any case, it seems that the particular arguments for the kind of ontological simplification I am arguing for is feasible for a variety of different meta-ontological positions. But I do not argue for this here, either.

[67] The reducing theory may not be true of course, hence the conditional.

ipso facto we have an empirically more adequate theory in the reducing theory.

Now, we may not be committed to the truth of the reducing theory - i.e. we might not think that the reducing theory itself is a good enough theory to include in our best conceptual scheme - but *if* it is then given the above, surely, the reduced theory is not. A way to throw this into a stark relief: suppose that the to-be-reduced theory is part of our best conceptual scheme and consider whether or not to include the reducing theory too. As we saw, the to-be-reduced theory cannot be used to explain the empirical success of the reducing theory.[68] Thus, one cannot 'get' the strength that the reducing theory would accrue to the conceptual scheme from the to-be-reduced theory. So either we have to forgo the strength that the reducing theory would accrue to the conceptual scheme or have both the to-be-reduced and the reducing theory as part of it. But neither disjunct is palatable it seems to me. We have every reason to want to include the reducing theory in our best conceptual scheme for it is more empirically successful than one we have already included! But then, if we do include it, why would we not 'kick-out' the to-be-reduced one? After all, we have an explanation of the empirical success of this theoy from the reducing theory. In short, if the to-be-reduced theory is good enough to include in our best conceptual scheme, then so is the reducing theory, but once the reducing theory *is* included, then the reducing-theory is redundant.

I concede that this is not an indefeasible argument. What makes it into our best conceptual scheme is, as Quine himself acknowledged, a matter of judgement. As such a definitive argument for the conclusion that a successful NN reduction entails ontological simplification is not forthcoming. But I do think that the above gives good reasons to think that it could. In any case, I think this is preferable to the untempered metaphysics one encounters in the literature.

1.5 Chapter Summary and Outlook

In this chapter, I have identified an important methodogical problem for theories of reduction, the *external* problem. I have solved it by taking a stipulative approach: abstract a general model of reduction from a rational reconstruction of the derivation of the Boyle Charles law the kinetic theory of gases. This formed the Neo-Nagelian model of intertheoretic reduction. The aim of reduction is ex-

[68] The requisite *warrant* would be lacking for the auxiliary assumptions if the explanation was 'reversed'. (cf. section 1.4.5)

planation. On this account, the reduction of one theory to another consists in deriving the laws of the former from the latter, and auxiliary assumptions and bridge-laws. However, there needs to be *warrant* for the auxiliary assumptions and bridge-laws for the derivation to be explanatory.

Before detailing the Neo-Nagelian model, I presented Nagel's model of reduction, and the amendments thereto. I showed why, even so amended, this model is not tenable. Its *internal* problems remain unsolved.

I then turned to the task of detailing and defending the Neo-Nagelian model. I proposed a general explanatory framework - *pluro-particularism* - and used this to rehabilitate the DN model of explanation which underpins the Neo-Nagelian model. With this in place, I explicated the notion of *warrant*. Finally, I also argued how successful Neo-Nagelian reduction affords ontological explanation. In so doing, I showed that, despite a structural similarity to the Nagelian model, the Neo-Nagelian model does not suffer from the *internal* problems besetting the former.

As aforementioned the derivation of the Boyle-Charles law was a mere sketch, and to substantiate the Neo-Nagelian model I presented here a closer examination of the derivation of the Boyle-Charles law is needed. This is undertaken in chapter 4.

However, first I shall show that so-called 'New Wave' models of reduction are not tenable alternatives to Nagel's model, as it is claimed. This forms chapter 2. Second, in chapter 3 I consider the issue of multiple realizability.

Chapter 2

New Wave Reductionism

2.1 Chapter 2 Introduction

In the previous chapter, I have set out the Neo-Nagelian model of reduction I am advocating. I have shown it to avoid the problems of the Nagelian model. However, purportedly there already is a superior model of reduction, the so-called 'New Wave' model. My aim in this chapter is to show that the so-called 'New Wave' model, in any of its guises, is not a tenable alternative to even Nagel's original model, and, *ispo facto* not a tenable altnernative to the Neo-Nagelian model either.

By the 1980s, Nagel's model of reduction was, despite various amendments, considered to be dead. Anti-reductionism became the dominant position, although developments in the sciences indicated that much research is done in a reductionist spirit. Neuroscience, for example, aims at an understanding of the human mind in terms of the microscopic constituents of the brain. Given the interest of the sciences in reduction and given that no acceptable model of reduction was forthcoming, philosophers, most notably Paul Churchland (Churchland, 1985) and Clifford Hooker (Hooker, 1981) proposed a new model of reduction. Purportedly, it avoids the problems associated with the Nagelian model, and is immune to the anti-reductionist arguments.[1] Following Hooker's and Churchland's work, the model has been developed by Bickle (1996, 1998). 'New Wave' reductionism centers on these philosophers' work.[2]

[1]Those advocating have been largely concerned with reduction, and reductionism in the context of philosophy of mind. Yet, the model is intended to be general and nothing hangs on their initial motivations.

[2]The term 'New Wave' is due to Bickle, and sometimes that term is used only to refer to his

New Wave Reduction, as I said, purportedly solves the problems pertaining to Nagel's model[3] and is putatively immune to the well-known anti-reductionist arguments based on multiple realizability. In the next chapter, chapter 3, I examine these anti-reductionist arguments and show that they have no purchase against Nagelian reduction and Neo-Nagelian reduction. Nonetheless, if the advocates of NWR are right, NN reduction is superfluous, for we already have a tenable alternative to Nagelian reduction.

In this chapter I shall argue that NWR reduction is not a tenable alternative to Nagelian reduction. NWR has been critically discussed in Eck et al. (2006), Schouten and de Jong (1998), Wright (2000) and most notably in Endicott (1998, 2001). Whilst I sympathize with these authors on the whole, I do not think their criticisms go far enough. In particular, the general thrust of these arguments has been that NWR is not such a 'new' model after all: on close inspection, NWR collapses into the Nagelian account it is claimed. This is only partially correct. Its advocates claim that a distinguishing feature of NWR is that it avoids recourse to bridge-laws. This is erroneous, and in this sense NWR does collapse back into the Nagelian model. However, I contend that NWR is in two important senses inferior to Nagelian reduction: first, it obfuscates important aspects of reduction and second, on the issue of ontological simplification, the position is hopelessly confused.[4] Arguing for this also serves a useful pedagogical purpose, for it will lend support to the Neo-Nagelian view of ontological simplification.

There is much similarity between the positions advocated by Churchland, Hooker and Bickle - hence the general term 'NWR' - but there are also differences, as shall become clear. I shall make a distinction between NWR as espoused by Churchland and the most recent incarnation owing to Bickle. The plan for this chapter is as follows. In section 2.2, I shall set out Churchland's 'informal' NWR model and show it to be problematic. In section 2.3 I shall set out the Hooker-Bickle model and show that it fares no better.

---

view. However, it is now more customary to use it to refer to the entire tradition which started with Churchland. In what follows I shall make the distinction between 'Churchland's Wave', which refers to Churchland's informal model of reduction, and 'Hooker-Bickle Wave' which refers to the formal view based on Hooker's work and developed by Bickle.

[3]These are what I have called the *internal* problems with Nagel's account. As set out in chapter 1 , the external problem for reduction is not acknowledged as a problem in the literature on reduction. For the purposes of this chapter I shall bracket this problem; I am here concerned with whether NWR is, as it claimed, a better alternative model of reduction.

[4]I also contend that Endicott is wrong in claiming that there *is* one novel feature in NWR (albeit a feature that is untenable, as he argues): the feature that Endicott singles out is not, to my mind, a novel feature at all. But this is a relatively minor point. cf. section 2.2.2.

2.2 Churchland's New Wave Model of Reduction

Paul Churchland proposes a new model of reduction in *Scientific Realism and the Plasticity of Mind* (Churchland, 1979). He also revisits reduction in his 1985 paper *Reduction, Qualia and the Direct Introspection of Brain States* (Churchland, 1985). In the 1985 work, he cites the 1979 work, as giving a more detailed account of the model being proposed.

I present the 1979 model followed by the 1985 model. Given that the earlier model is said to be a more detailed version of, and hence congruent with, the latter, I shall concentrate my criticism on a synthesis of the two. The selling point of his model is that it is "more accurate, general and illuminating [than the Nagelian model]" (Churchland, 1985, 10). I argue that it is not.

2.2.1 Churchland's Wave

Churchland considers two theories, an old and a new, denoted by T_O and T_N.[5] These are the 'to-be-reduced' theory and the 'reducing theory' respectively. He then states two desiderata for a successful reduction:

> (CD1) "[I]t provides us with a set of rules – "correspondence rules" or " bridge laws" [...] – which effect a mapping of the terms of the old theory (T_O) onto a subset of the expressions of the new or reducing theory (T_N) ..." (Churchland, 1979, 81-82)
>
> (CD2) "[A] successful reduction ideally has the outcome that, under the term mapping effected by the correspondence rules, the central principles of T_O [...] are mapped onto general sentences of T_N that are *theorems* of T_N. Call the set of such sentences S_N. This set is the image of T_O within T_N... [In this sense,] T_N contains as a proper substructure, a set of principles S_N that is isomorphic with the set of principles comprising T_O." (*ibid.*)

With these desiderata in place, Churchland makes his central claim:

> "*successful reduction is a fell-swoop proof of displaceability*; and it succeeds by showing that the new theory contains as a substructure an equipotent image of the old." (Churchland, 1979, 82, orig. emph.)

[5] Notice that using the labels 'old' and 'new' suggests that what Churchland has in mind is diachronic intertheoretic reduction. However, his discussion and intended application vis. the relation between folk psychology and neuroscience, say, indicate that the model is not to be so restricted.

This, continues Churchland, is an 'ideal or maximally smooth' (*ibid.*) case however. In general the situation will be more involved in two ways. First, the image of T_O within T_N may not be a direct consequence T_N *alone*; it may be deducible only from an 'augmented theory', $T_N^{'}$, comprising of T_N plus auxiliary assumptions. (cf. *ibid.*, 83-84). Second, we may not be able to derive an exact image of T_O within $T_N^{'}$, and may have to rest content with deriving a modified or corrected version $T_O^{'}$ of T_O. In this case we refer to S_N as the 'corrected image of T_O'. We then require that the corrected theory $T_O^{'}$ and the original theory T_O be "*closely similar*" (*ibid.*, 83). This exhausts the content of Churchland's 1979 model of reduction.

In the 1985 version, we once more start with an old and a new theory, T_O and T_N. (Again, these are the 'to-be-reduced' theory and the 'reducing theory' respectively.) Churchland then offers the following schema:

> "T_N & Limiting Assumptions & Boundary Conditions
>
> logically entails
>
> I_N [a set of theorems of (restricted) T_N]
>
> e.g. $(x)\ (Ax \supset Bx)$
>
> $(x)\ ((Bx\&Cx) \supset Dx)$
>
> which is relevantly isomorphic with
>
> T_O (the older theory)
>
> e.g. $(x)\ (Jx \supset Kx)$
>
> $(x)\ ((Kx\&Lx) \supset Mx)$"
>
> Churchland (1985, 8)

Churchland expands on the schema; it is worth following the text closely:

> "[A] reduction consists in the deduction, within T_N, not of T_O itself, but rather of a roughly equipotent *image* of T_O, an image still expressed in the vocabulary proper to T_N. The correspondence rules play no part whatever in the *de*duction. They show up only later, and not necessarily as material-mode [identity] statements, but as mere ordered pairs: $< Ax, Jx >$, $< Bx, Kx >$, $< Cx, Lx >$, $< Dx,$

> $Mx >$. Their function is to indicate which term substitutions in the image I_N will yield the principles of T_O. The older theory, accordingly, is never deduced; it is just the target of a relevantly adequate *mimicry*. Churchland (1985, 10-11, orig. emph.)

As already noted, Churchland's 1985 paper directs us to the 1979 paper for a more detailed exposition of the model. Thus, I shall take seeming conflicts between the papers to be mere differences in emphasis or style and I shall, in presenting my criticisms, direct them towards, what I hope is, the most charitable synthesis of the two papers, as follows:[6]

> Churchland's NWR: From T_N, limiting assumptions and boundary conditions, we deduce a set of theorems, call them I_N.[7] We then substitute the predicates in this set for predicates from the to-be-reduced theory, using the list of ordered-pairs, e.g. $< Ax, Jx >, < Bx, Kx >, < Cx, Lx >, \langle Dx, Mx \rangle$. In the ideal case, this will yield T_O, in which case I_N will be said to be an 'equipotent image' of T_O. However, it may turn out that we do not get an *exactly* equipotent image of T_O, but something 'closely similar' to it (or 'roughly isomorphic' to it), where this is denoted by $T_O^{'}$ (That is, we deduce an 'almost equipotent' or 'roughly isomorphic' image of the to-be-reduced theory, $I_N^{'}$ and substitute the relevant predicates as per the ordered pairs to yield $T_O^{'}$).

2.2.2 Problems with Churchland's Wave

As aforementioned, the general thrust of the criticism leveled against NWR is that it is Nagelian reduction in disguise. I think that this is only *partially* correct. Whilst there are similarities between the two, NWR *is* distinct from the Nagelian account in certain respects. However, where they differ, NWR comes off worse, or so I shall argue.

Let me first articulate the alledged differences between NWR and the Nagelian model. The first obvious difference between the two models is the language. Churchland's model makes use of set-theoretic or model-theoretic language. A second alleged difference (although this is a consequence of the first) is that the

[6] I set aside the incredibly idealised case where recourse to limiting assumptions and boundary conditions is not necessary.

[7] Throughout the text the subscripts indicate which of the two theories' vocabularies the subscripted term is couched in.

to-be-reduced theory is never deduced but the "target of a relevantly adequate *mimicry*" Churchland (1985, 11). Third, and most importantly, Churchland's NWR allegedly avoids bridge-laws.

Churchland's NWR makes use of model-theoretic language. There are 'sets', 'mappings', 'isomorphisms', 'images' and so forth. Clearly this alludes to the semantic view of theories, initiated by Suppes (1960) and since developed. (cf. for example, Suppe (1974), Fraassen (1980), and Giere (1994)). NWR as advocated and developed by Bickle (1996) offers a detailed model-theoretic model of reduction based on the structuralist program of philosophy of science associated with the 'Munich School'. This is discussed in section 2.3. In Churchland's case, however, the allusion seems to be just that: an allusion. Certainly, the notions of 'relevant isomorphism' and images being 'roughly equipotent' are simply not part of the formal model-theoretic language of the 'Munich School'. This alone suggests that Churchland intended this is to be informal.[8] It is easy to see why critics of NWR have argued that this basically the Nagelian model with a structuralist veneer. But, at least in this respect, I think the right thing to say is that Churchland's account is, in fact, too vague to really decide the matter. What it is for two models to be relevantly isomorphic to one another? Even granting that in the ideal case, 'isomorphic' is used colloquially to mean two models having identical structures, what sense of 'relevancy' is in play here? Churchland does not explain this. Perhaps what is intended is that there is an isomorphism between certain parts of the models - the 'relevant' parts - but again what those parts would be, is left unarticulated. In the non-ideal case, we are told that what must be deduced within T_N is a corrected image of T_O, $T_O^{'}$. Yet no indication is given as to what this similarity relation consists of. The charge is not that Churchland does not give some formal abstract measure of 'similarity' here - as I argued for NN reduction in the previous chapter, I do not think that it is fruitful to give some such measure - but that he does not (even) indicate any *informal* criteria for what counts as 'similarity'. In short, the problem is not that Churchland's NWR is not genuinely distinct from Nagel's model but that it is too vague to be a tenable alternative, at least in this respect.

The second alleged difference is really only a consequence of the first, but it is useful to consider it separately, for it is *putatively* a *distinct* upshot of the Churchland's NWR. Churchland claims that that the to-be-reduced theory is

[8]If it is intended to be formal it can be subsumed into Bickle's account discussed below.

never deduced but is only the "target of a relevantly adequate *mimicry*" (*op. cit.*) avoids one of the problems of Nagel's model, namely the Falsity problem (cf. chapter 1.3.2.2). Recall this was the problem of deriving a false theory from a (supposedly) true one. I argued that this problem dissolves once we note that the auxiliary assumptions are false and so this is not a problem for Nagelian reduction (nor, indeed, for NN reduction). But suppose that this were an outstanding problem for Nagelian reduction. Is Churchland's solution to the problem persuasive? Churchland claims that the to-be-reduced theory is never deduced and hence that the Falsity problem is a non-starter on the NWR account. But this is not right: in the ideal case where we deduce an equipotent image, I_N, of T_O from T_N, we *do* then 'deduce' T_O too, simply by the substitutions given by the ordered-pairs, $< Ax, Jx >, \ldots$! Now, strictly speaking, set-theoretic *replacement* is not the same thing as *deduction*. But surely nothing of substance hangs on this. So this is not a salient difference.

This brings us neatly to the final alleged difference between NWR and Nagelian reduction: the lack of bridge-laws. It is obvious that the 'mere' ordered-pairs serve the same *function* as bridge-laws in Nagel's account. This much is not in dispute but it is worthwhile articulating why it is that (at least something like) bridge-laws are part of Churchland's NWR. It won't do for I_N and T_O to only be *formally* similar: two models can be *formally* identical and yet be wholly unrelated. For example, certain equations in macroeconomics are formally identical to equations in hydrodynamics[9] but of course we do not want to regard the former as reducing to the latter! This would be a, so-called, *spurious* reduction.[10] To avoid spurious reduction bridge-laws are needed. I am in agreement with Endicott when he pre-empts a possible response:

> Nor will it do, as a response, to insist on a distinction between the "reduction proper" versus its "consequences," confining bridge-laws to the latter. Consequences are consequences, and to deny them is like a smuggler caught in the act whose only defense is: "I meant there was no contraband *on my person*!" Declared or no [*sic.*], up front or

[9]This is no accident. The macroeconomic equations were purposefully modeled on the hydrodynamics ones. cf. Mirowski (1991)

[10]That similarity (or identity) of structure is not sufficient to ground a reduction is an argument that is put forward against Suppes' model of reduction (Suppes, 1960) by Schaffner (1967). Bickle characterises this as a general problem for structuralist models of reduction - the "too weak to be adequate" challenge (Bickle, 1996, 74). I do not consider Suppes' model separately, as all of its main features are subsumed into Bickle's account. c.f. section 2.3.

trailing behind in tow, the goods are there..." (Endicott, 1998, 69-70)

That there are bridge-laws, in all but name, in Churchland's NWR also ushers in the same problems as those we saw for bridge-laws on Nagel's account: Where-Hence and Status. Recall that these are the problems of determining the proper form of bridge-laws (i.e. why it is that, say, temperature is associated with mean kinetic energy, as opposed to, say mean kinetic energy squared, or particles center of mass, etc.) and what their status is. Churchland has nothing to say about the former; this problem is as much a problem for Churchland's NWR as it is for the Nagelian account, Bridge-laws (or the list of ordered-pairs) are just taken as a given. As regards the latter, however, Churchland provides a much touted solution. This is that the status of bridge-laws is determined in light of the reduction itself.

Churchland contends that one can use bridge-laws *qua* 'mere' ordered pairs in the reduction and that they are only interpreted post-reduction. This idea is untenable: to give a reason for choosing a particular bridge-law requires them to be more than mere ordered pairs; one cannot motivate a putative bridge-laws if it is literally construed as indicating the substitution of symbols.[11] To put the point in terms of the bridge-law in the derivation of the Boyle-Charles law, it is not 'temperature' which is associated with 'mean kinetic energy' but rather *it is temperature that is associated with mean kinetic energy.* Of course, one may wish to be agnostic about just what the status of this is - i.e. remain agnostic about the interpretation of 'associate' - but that is different from saying that all that is involved is the substitution of symbols.

But suppose that God slides a list of the correct ordered-pairs into your hands. You know for sure which predicates to 'associate' with which. With this to hand, you set about the task of trying to determine their status.

Churchland's idea is that the status of the bridge-laws is determined by the comparative *smoothness* of the relevant reduction. He considers a "smooth" reduction to be one where the limiting assumptions and boundary conditions are not "wildly counterfactual" and where "all or most of the principles of T_O find close analogies in I_N, etc" (Churchland, 1985, 11). Specifically:

"[S]moothness permits the comfortable assimilation of the old ontol-

[11] In the terminology of NN reduction, one needs *warrant* for bridge-laws, which, amongst other things, consists in arguing for *conceptual fit* between the associated properties in the bridge-laws. This cannot be done if the bridge-laws are taken to be mere substitutions of symbols.

> ogy within the new and thus allows the old theory to retain all or most of its ontological integrity. *It is the smooth intertheoretic reductions that motivate and sustain statements of cross-theoretic identity, not the other way around.*" (Churchland, 1985, 11 orig. emph.)

Why is it that a 'smooth' reduction allows the ontology of the to-be-reduced to be 'comfortably assimilated' by the reducing theory? Why does a 'smooth' reduction license an identity claim? Churchland simply does not provide an answer to these questions. Let me use possible worlds semantics to shed light on the issue. The intuition underpinning Churchland's claim seems to be that the 'closeness' of the possible world at which the to-be-reduced theory is true, PW_{TBR}, to the possible world at which the reducing theory is true, PW_R, is what determines whether or not one identifies the ontology of the two theories.[12] But this is problematic in several respects. First, the 'closeness' of possible worlds in this sense is a matter of degree - the auxiliary assumptions can be *more or less* counterfactual - which is incongruent with the identity being a binary relation. Even other problems notwithstanding, surely deciding where one has identity and where not on the counterfactual spectrum is going to be arbitrary. Second, it is not clear how the bridge-law is to be interpreted when PW_R and PW_{TBR} are not 'close'. Would a less-than-smooth reduction justify the postulation of, say, a nomic connection between the properties instead? Third, even bracketing these two problems, why think that the 'closeness' of PW_R and PW_{TBR} gives licence to an identity claim in the first place? Suppose for the sake of argument that the empirical adequacy of a theory is a good guide as to whether or not the properties it posits refer to actual properties. And suppose further that PW_R is 'close enough' to the actual world, PW_A, to justify the claim that its properties do refer. Then you might think, that if PW_{TBR} is very close to PW_R, then this is good reason to think that the properties that PW_{TBR} posits also refer to actual properties.[13] But even if so, this does not entail that the PW_R and PW_{TBR} refer to the same properties. They may refer to different properties which are correlated (nomically or otherwise). So, even granting everything else, the 'closeness' of PW_R and PW_{TBR} *per se* does not establish the identity of the properties of

[12] Notice that what is in the offing here is the identity of the properties of the two theories, not their entities, via bridge-laws.

[13] Now, I think that this is muddled thinking. My position on ontological simplification is set out in chapter 1.4.7 but this is not my concern here.

the two theories.[14] In contrast to this, I claim that a successful reduction of one theory to another is good grounds to be ontologically uncommitted to the former. For, we have an explanation of the empirical success of the to-be-reduced theory via the supposition of the reducing theory, which, *ex hypothesi*, is an empirically more adequate one. That is, we have explained why (to the extent that the to-be-reduced theory is empirically adequate) it is *as if* the to-be-reduced theory is true without it *actually* being so.

So where do we stand? As I have said, the thrust of the criticisms of NWR has been that it is essentially Nagel's model disguised in a structuralist veneer. This is partially right, for Churchland's NWR bares some of the hallmarks of Nagelian reduction, viz. the derivation of the laws of the to-be-reduced theory and use of bridge-laws, in all but name. However, in other respects Churchland's NWR is actually worse: the emphasis on 'mimicry' of the to-be-reduced theory rather than the 'deduction' of it, and the colloquial use of structuralist language are, at best, stylistic amendments which tend to obfuscate some of the important issues here. Finally, I considered Churchland's NWR claim that 'cross-theoretic' identities can be established post-reduction dependent on the counterfactualness of the auxiliary assumptions. Churchland does not provide an argument for this position. However, I have argued that it seems an untenable one in any case.

Before moving on to consider Hooker-Bickle NWR, I wish to consider for completeness a point raised by Endicott against NWR. Endicott argues that there is one genuinely novel feature of NWR[15], namely:

> "New-wave construction: the basic reducing T_N not the original reduced T_O supplies the conceptual resources for constructing the corrected T'_O... [It is only this condition that] is the genuinely novel element: the basic reducing T_N and not the original reduced T_O must supply the conceptual resources for the corrected image T'_O." (Endicott, 1998, 56)

Endicott goes on to argue that this feature is not ultimately defensible. Be that as it may, it is not clear to me in why it is that Endicott considers it to

[14] In chapter 4 I argue that bridge-laws *qua* identities are in fact entirely misguided but I bracket this issue here.

[15] Endicott synthesizes both Churchland's and Hooker's informal models with Bickle's formal one but, whilst useful in that it allows Endicott to criticise all three advocates at once, glosses over distinctions between them (albeit differences that are ultimately not defensible). I have changed the subscripts in the following quotation simply to make it congruent with the above.

be a genuinely novel feature. Unfortunately, Endicott does not detail why he considers it to be so. *Prima facie*, it seems that this is a trivial condition of the (Schaffner-modified) Nagelian model: certainly on that account T_O *does not supply* the conceptual resources for *constructing* $T_O^{'}$; $T_O^{'}$ is to be *compared* to T_O. There are several reasons one *might* consider condition (i) to be a genuinely novel condition but, I argue, upon closer examination we see that it is not. 1) The emphasis might be on 'construct': as opposed to the Nagelian model where the emphasis is on deduction, here, the thought might be, one is constructing $T_O^{'}$ from T_N and the auxiliary assumptions. But, I think rather obviously, this would not be a substantial difference but one of mere expression, as we have seen - another instance of the structuralist veneer. 2) Another reading would that $T_O^{'}$ is constructed (derived) using only the conceptual resources of T_N in the sense that $T_O^{'}$ is couched in terms of the properties and/or ontology of the former. To echo Churchland's claims, the idea is that T_O is never deduced but is, rather, mimicked. But again, we have seen that this mimicry is also just another difference in gloss.

2.3 The Hooker-Bickle New Wave Model of Reduction

In the previous section, Churchland's model was critically discussed; in this section I consider Hooker and Bickle together. It is preferable to consider them together rather than individually because Hooker's original model devoid of Bickle's thorough-going structuralist reworking of it, is not substantially different to Churchland's.

Bickle gives much kudos to Hooker's insights (the second chapter of six of *Psychoneural Reduction The New Wave* is entitled 'Exploiting Hooker's Insights'). I consider 'Hooker's insight' first in section 2.3.1.

Bickle's self-stated goal is to provide a theory of reduction "which exploits some of Clifford Hooker's insights about the nature and consequences of scientific reduction" (Bickle, 1996, 57) but one which avoids "a limitation in Hooker's general account" (*ibid.* 58) namely that the 'analog relation' is not elucidated. To do this Bickle seeks to:

> "embed Hooker's insights within an approach to the structure of scientific theories and intertheoretic relations that at least provides a

start on a precise, formal account of the implied amount of correction to the reduced theory." (*ibid.* 58)

I consider Bickle's formal model in section 2.3.2. I show it to be untenable.

2.3.1 Hooker's Insights

Bickle frames Hooker's insights in historical context, namely as appearing as "a response to well-known criticisms of the "received-view" stemming from Ernest Nagel and logical empiricism." (Bickle 1998 23) In particular, the problem that Hooker is concerned with[16] is that of deducing a supposedly false theory from a supposedly true one. (This is the 'Falsity' problem as per chapter 1.3.2.2.)

As Bickle has it, there are two distinct possible responses to this problem: either "one can supplement the reducing complex (the premises of the deduction) with various boundary conditions and limiting conditions, some counter to fact." (Bickle, 1996, 24) or one can go the 'Schaffnerian route' in which the to-be-reduced theory is not deduced but instead a 'corrected version' of the to-be-reduced is deduced. In the former, the logical problem is obviously avoided because the false to-be-reduced theory is deduced from a set of premises some of which (viz. the auxiliary assumptions) are false. In the latter, Bickle's thought seems to be what is deduced is a 'corrected' theory, in the sense that it is (by supposition) true, thus avoiding the logical problem.

Bickle's disjunction is misleading. Auxiliary assumptions are always needed and what Bickle calls the 'Schaffnerian' route was never intended as a solution to the 'Falsity' problem. Schaffner was concerned with the case in which, *even given certain auxiliary assumptions*, one cannot derive the exact laws of the to-be-reduced theory. Indeed, the example considered by Schaffner is the thermodynamics to statistical mechanics case, for which counterfactual auxiliary assumptions are needed (by Schaffner's lights, and correctly so).

Be that as it may, Bickle is concerned that:

> "[n]either of these approaches handles every problem raised by reductions of empirically false theories. Sometimes a theory reducible to (some portion of) its successor turns out to be so radically false (in certain respects) that central elements of its ontology must be rejected as empirically uninstantiated." (Bickle, 1996, 25)

[16] At least, according to Bickle.

This leads to a problem that neither of the above 'solutions' (or 'approaches') can handle, namely:

> "[the] referents of terms of [the reducing theory] cannot be synthetically identical to, or even nomically coextensive with, "referents" of terms of [the to-be-reduced theory or the 'correct version' of it] if the latter completely lack actual extension." (*ibid.*)

Note the slide from the question of how the laws of false theory can be derived from a true theory to a worry about ontology. These are different problems! But the confusions do not end there. Bickle proffers the following diagnosis of the latter problem:

> "The problematic cases arise from their treating reduction not just as deduction but as deduction of a *structure specified within the vocabulary and framework of the reduced theory*–either the [to-be-reduced theory] itself or some correct version [of it]." (Bickle, 1996, 26 orig. emph.)

With this diagnosis in place, the solution readily presents itself: give up this feature altogether! And giving up this feature is what Bickle takes to be Hooker's first important insight about intertheoretic reduction.

> "On Hooker's account, neither [the to-be-reduced theory] itself nor any structure constructed from its vocabulary and explanatory resources gets deduced in a reduction, not even in the smoothest cases. [Rather the to-be-reduced] is always the target of a kind of *complex mimicry*..." (Bickle, 1996, 27 orig. emph.)

Of course, this sounds familiar - it is just like the Churchland account discussed above[17] but it is worth revisiting all this because, first, Bickle expands on "Churchland's schematic illustration of Hooker's account" (ibid) and it is worth considering whether thus expanded upon the view is any more tenable. Second, it allows us to critically discuss the formal model of reduction he (Bickle) develops off the back of Hooker's (and Churchland's) without fear of having misrepresented him.

It is pertinent to consider the passage in which Bickle explains how this "single change addresses the 'radical falsity' worry" (Bickle, 1996, 26 orig. emph.) in full.

[17] Indeed, Bickle cites and quotes Churchland in these passages.

> "Neither connecting principles nor reductive [bridge-laws] play any role in the derivation of Hooker's T_O^*. There is no need for them. T_O^* is already specified within (a restricted portion of) the vocabulary of T_B. Elements analogous in some respects to connecting principles occur within Hooker's account, when we explore both the nature of the analog relation AR obtaining between T_O^* and T_O and the ontological consequences of a given reduction. But these elements are merely ordered pairs of terms, drawn from the nonlogical vocabularies of the two theories. Their sole function is to indicate the term substitutions in T_O^* that will yield the laws of T_O (or approximations of those laws, depending upon the extent to which the reduction corrects T_O). No worry arises about the "logical status" of ordered pairs of terms, even when one of a pair has no empirical extension. By themselves, these ordered pairs imply neither synonymy, synthetic identity, nor coextension." (Bickle, 1996, 28)[18]

I think that this is not persuasive for the following reasons. First, stressing that the derivation of T_O^* does not involve bridge-laws at all is simply a matter of terminology. Bickle can resist calling the ordered pairs 'bridge-laws' but this is nothing substantive. The ordered-pairs play the same function as bridge-laws. Second, as with Churchland's NWR, the claim that they come after the 'reduction proper' cuts no mustard either; recall Endicott's smuggler. Third, and again as with Churchland's NWR, the suggestion that the bridge-laws are mere ordered pairs of terms is untenable: any reason for associating two terms requires that the terms be understood as more than just syntactic elements (or labels). And some reason, roughly speaking, there must be for not any old association will do on pain of the spurious reduction problem. (c.f. section 2.2.2.) In short, Bickle's NWR, like Churchland's, does not avoid recourse to bridge-laws and hence does not avoid the problems associated with them. Like Churchland, Bickle also endorses (in fact he attributes this to Hooker) the idea that post-reduction, the smoothness of a reduction justifies identity claims for bridge-laws. But Bickle at least purports to extend this proposal. To this I now turn.

Suppose that God slides Bickle a list of the bridge-laws too. Bickle contends that 'smooth' reductions justify bridge-laws qua identities, as with Churchland.

[18]Note that by 'connecting principles' Bickle is referring to Churchland's ordered-pairs, as per section 2.2.1.

Unlike Churchland, however, (who is mute on this), Bickle proposes that when the reduction is not smooth this justifies an elimination of the property or entity involved in the given bridge law:

> "Relatively smooth intertheoretic reductions are affiliated with robustly retentive ontological consequences, while relatively bumpy reductions are affiliated with eliminativist conclusions." (Bickle, 1996, 31)

Consider the diagram that Bickle provides to summarize his position, Figure 2.1.

The top line indicates the 'intertheoretic reduction spectrum', from 'smooth' on the left to 'bumpy' on the right. The bottom line indicates the 'ontology spectrum': at the bottom the far left there is identity and on the far right there is elimination. Along the top line there are various pairs of theories – the pairs that putatively stand in reductive relations of varying smoothness (bumpiness) to each other. As example, we are given physical optics to the electromagnetic theory as a putative 'smooth' reduction (on the far left) and phlogiston chemistry to oxygen chemistry as a putative 'bumpy' reduction (on the far right). The in-between cases are Keplerian to Newtonian mechanics and thermodynamics and statistical mechanics. The vertical arrows run to putative bridge-laws in these cases. So for the optics case, we get "visible light = EM radiation..." and the chemistry case, 'phlogiston does not exist...'. In the 'in-between' cases the arrows run to 'planetary orbits not quite elliptical' and 'ideal gas (strictly) = SM constructs only in mathematical limit'.

There are several very baffling elements to this diagram: First, the vertical two-ended arrows, that are labeled 'rough isomorphism'. These run from the *pairs* of the theories towards the 'ontology spectrum' (the bottom line). Recall that a rough isomorphism is something that is supposed to hold between two theories - it is thermodynamics that is supposed to be roughly isomorphic with statistical mechanics, for example. In what sense can there be a rough isomorphism between pairs of theories and bridge-laws?! Second, the putative bridge-laws in the 'in-between' cases are not bridge-laws at all: that the orbits of planets are not quite elliptical or that the strict ideal gas law can only be obtained from statistical mechanics in some mathematical limit are not bridge-laws! Third, even the putative bridge-laws at the two ends of the spectrum are odd: the bridge-laws that

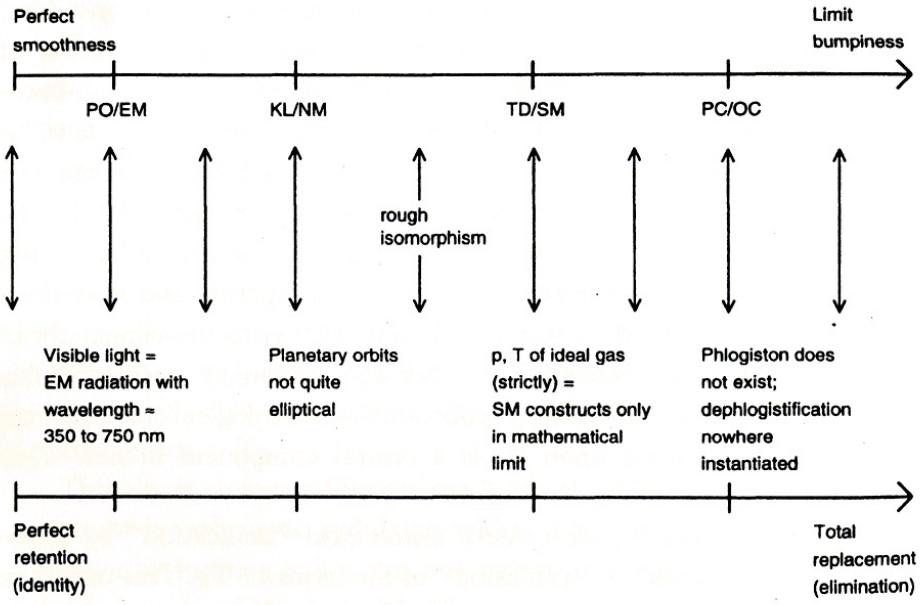

Figure 2.1
Top arrow: the intertheoretic reduction spectrum. Some historical cases are ranked (ordinally) according to the amount of correction to the reduced theory. Bottom arrow: the ontology spectrum. Cross-theoretic ontological consequences affiliated with these historical cases are ranked (ordinally) from retentive to eliminative. PO: physical optics (wave theory of light); EM: Maxwell's electromagnetic theory; KL: Kepler's laws of planetary motion; NM: Newtonian mechanics; TD: classical equilibrium thermodynamics; SM: statistical mechanics (kinetic/corpuscular theory of heat); PC: phlogiston chemistry; OC: oxygen chemistry. (From Bickle 1996.)

Figure 2.1: Bickle's formal model (Bickle 1998 31)

Bickle was hitherto concerned with connected properties not entities. Now there may be entity bridge-laws too, but nowhere do we see a bridge-law connecting properties, like the temperature-mean kinetic energy bridge law.

Even setting these problems aside, one gets no insight into why a 'smooth' reduction affords identity and a 'bumpy' reduction elimination. Bickle says nothing whatsoever about this other than "[c]onsequences grow respectively more revisionary and finally eliminative for cases falling toward the [bumpy limit]." (Bickle, 1996, 31) The very same problems that beset Churchland's NWR apropos this issue apply to Bickle's. (cf. section 2.2.2.)

Let us try to take stock at this point. What were the supposedly new aspects of New Wave Reduction and in what way were they supposed to be an improvement to Nagel's model? I argued that the colloquial use of structuralist language, far from improving on Nagel's model, just obfuscates the important aspects of reduction. In terms of substance, the two concrete differences encountered so far between Bickle's NWR and the Nagelian model were supposed to be that, first, one can circumvent the 'radical falsity' worry by avoiding having to specify the status of bridge laws in the reduction itself (the bridge laws, it was claimed, are mere ordered pairs) and, second, that one determines the status of the bridge laws *after* the reduction has taken place, in virtue of the smoothness of the reduction itself. I have argued against the very coherence of both suggestions.

As I said right at the beginning of this section, Hooker's model of reduction is very similar to Churchland's and it is therefore not surprising that I have criticized them in much the same way. Still, I suggested that it was worthwhile to go through Bickle's rendition of Hooker's model for two reasons. First, Bickle provides far more by way of detail and defence (though ultimately in vain, I argued) to the informal model and it was important to address this. Second, Bickle uses Hooker's model as platform for developing a *formal* model of reduction in the structuralist mold. It is doubtful that taking a flawed informal model and formalising it will, as it were, make it any more tenable but let us not be hasty.

Bickle of course does not recognise the flaws in the informal model. Suppose, *pace* my arguments above, that the informal model is tenable. An obvious question to ask is why then go down the formal route at all? Bickle's reason for going down the formal path is the following: he claims that a thorough-going structuralist rendition of reduction allows him to address an incompleteness in Hooker's original model (and in fact all the other informal models of reduction), namely

that in doing so he can make precise the spectrum of intertheoretic reductions. But why is this desirable?

> "Here some sympathizers to new-wave reduction will demur. Some question any need to find a formally specifiable measure for the intertheoretic reduction spectrum" (Bickle, 1996, 54)[19]

However, Bickle thinks that this would be inadvisable:

> "[W]e should not ignore the disastrous limitations of this attitude if we wish to draw ontological conclusions based on the nature of the intertheoretic-reduction relation." (*ibid.*)

Bickle's point is that, given that the smoothness or bumpiness of a reduction determines (ought to determine) one's ontological commitments post-reduction, it is only with a formal measure of the smoothness that this can be made precise. However, I have argued extensively that this link between the smoothness of a reduction and ontologoy is ill-conceived. Whilst I think it doubtful that formalising the model will make it any more tenable, not to consider it all would be too hasty, as I said. In the next section I present Bickle's formal model and show that none of the problems that beset the informal new wave model(s) of reduction are thereby solved (as was to be expected!).

2.3.2 Bickle's Formal Model

Bickle proffers a theory of reduction based on the "structuralist philosophy of science" of Balzer, Moulines, and Sneed (Balzer et al., 1987).[20] I closely follow Bickle's exposition of that position in what follows.

First some stage-setting. Bickle suggests two characteristics of "semantic views" in general:

> (B1) that scientific theories "are properly conceived *not* primarily as linguistic entities (i.e., sets of sentences), but instead in terms of their *models*." (Bickle, 1996, 59)

[19] In particular he cites Patricia Churchland as someone who argues that there is no need for formal criteria that allow one to place a given reduction on the smooth-to-bumpy spectrum. cf. Churchland (1986).

[20] Bickle's approach falls under the umbrella of the "semantic view of theories". Bickle arguably takes his cue from the work of Suppes, in that Suppes' was the first to put forward a theory of reduction based on the, then emerging, "semantic view". (cf. Suppes (1960)) Unless otherwise specified, in what follows by 'structuralism' I refer to the structuralism attributed to Balzer et al by Bickle.

(B2) that the "appropriate tool for the formal explication of the structure of scientific theories is not first-order logic and metamathematics, but instead *mathematics*." (*ibid.*).

The claim that one explicates the *structure* of scientific theories in mathematics is somewhat misleading for one explicates the *content* of scientific theories as well as the structure in this way. Indeed, as Bickle himself puts it: "Models here are not representations of things depicted by a theory. Instead they [the models] *are* the things depicted." (Bickle, 1996, 62) Or again: "Instead of saying that certain equations are a model of subatomic or economic phenomena, we propose to say that the subatomic or economic phenomena are models of the theory presented by those equations" (Balzer et al., 1987, 2))

Consider now the structuralist machinery. The language of structuralism is set theory, and to give a structuralist rendition of a particular theory one starts with a set-theoretic predicate "is a model of T". Something is a model of a theory T if it has some specified formal properties. There are three important notions here:

(PM) *Potential* models of T, $M_p(T)$

(AM) *Actual* models of T, $M(T)$

(IEA) *Intended empirical applications* of T, $I(T)$

The *potential* models of T, $M_p(T)$, satisfy certain formal properties. We can think of the potential models as forming an abstract conceptual framework. The *actual* models of a theory T, $M(T)$, are a subset of the $M_p(T)$, i.e. $M(T) \subseteq M_p(T)$. These are models which satisfy further law-like properties. Thirdly, there are *intended empirical applications* of the theory, $I(T)$, which are also a subset of $M_p(T)$, i.e. $I(T) \subseteq M_p(T)$. Bickles continues:

> "seemingly without exception $M(T) \subset M_p(T)$... and typically at any given time $I(T) \cap M(T) \neq \emptyset$ but $I(T) \not\subseteq M(T)$." (Bickle, 1996, 64)[21]

Now consider two theories that have been given a structuralist rendition. Bickle notes that structuralists have proffered various accounts of intertheoretic reduction. Formally, one characterises reduction by placing constraints on the

[21] cf. (Bickle, 1996, 65) for why $I(T) \not\subseteq M(T)$.

so-called reduction relation, ρ, where ρ is a relation between the (model-theoretic rendition of the) physical theories. Bickle contends, however, that the constraints placed on ρ do not "capture Hooker's insight about what gets deduced in a reduction" (Bickle, 1996, 65), namely an analog structure T_O^*, rather than T_O itself, "specified within the framework of the reducing theory T_B and designed to mimic the structure of T_O." (*ibid.*)

Bickle proposes to capture 'Hooker's insight' within the basic structuralist framework by offering a different characterisation of the reduction relation – ρ – to those already offered by the structuralists. Again, 'Hooker's insight' is the starting point:

> "I begin with Hooker's notion of the analog structure T_O^*. T_O^* is a structure specified within a restricted portion of T_B that is analogous to T_O." (Bickle, 1996, 65)

One then adds, on this view, the auxiliary assumptions as axioms of the relevant predicates characterizing the reducing theory. Bickle notes that if any of these axioms are counterfactual then T_O^* cannot be a proper subset of (the models of) T_B. What of Hooker's analogy relation, "AR" or, as referred to here, "ρ"? Bickle firstly considers the two constraints put forward by Balzer et al. These are as follows:

> (BMS1) $\forall x' \; \forall x, x' \in T_O^*, \; x \in M_p(T_O)$: if $x' \; \in M(T_B)$ and $< x', x > \in \rho$, then $x \; \in M(T_O)$.
>
> (BMS2) ρ "relates some elements of T_O^* to merely potential model of T_O" (Bickle, 1996, 69)
>
> (BMS3) ρ "relate a *confirmed* intended empirical application of T_B—a "real world" potential model in $I(T_B)$ already shown to be an actual model—also belonging to T_O^* to each confirmed intended empirical application of T_O." (*ibid.* 70)

BMS1 requires that ρ only relates:

> "actual models of the T_O [the to-be reduced theory] to actual models of T_B [the reducing theory] that also belong to T_O^*, i.e., that ρ never relate an actual model of T_B also contained in T_O^* to a merely potential model of T_O." (Bickle, 1996, 68)

The conjunction of these constraints entails the formal analog to Nagel's derivability condition. (cf. (Bickle, 1996, 69) and (Balzer et al., 1987, 275)) These, he suggests,

> "are possibly too strong to account for all of the actual historical cases that scientists have dubbed "reductions"... Hooker denies that his relation AR between T_O^* and T_O needs to be as strong as derivability, even in the smoothest cases of reduction." (*ibid.*)[22]

He goes on, therefore, to consider a weaker set of conditions, as per Mayr (1976) and formalises this as a further constraint on ρ. The details need not concern us for Bickle thinks that it too is effectively too strong to deal with every actual case of reduction. Indeed, it fails to deal with the relation of simple thermodynamics to kinetic theory which is an instance of reduction *par excellence* for Bickle. Hence, Bickle proposes an even weaker set of constraints. Again the details can be glossed over for the substantial point comes next. Bickle states:

> "Even with all this machinery I still can't hint at how to reconstruct the one historical case of reduction central to the new-wave reductionist program: that of simple thermodynamics of gases to the kinetic theory and statistical mechanics." (Bickle, 1996, 73)

That is, even with *his* further weakening of the formal constraint on ρ, the NWR account he advocates cannot account for this case. Why? Because, as it stands, Bickle's formal account faces a far more pressing problem. This is the "spurious reduction" problem. (cf. section 2.2.2.) As Bickles notes, the conditions on ρ may very well be satisfied between models of theories that are wholly distinct and obviously not reductively related. For example, we might, without imposing further requirements, absurdly conclude that certain parts of macroeconomics reduce to hydrodynamics. Bickle concedes, that:

> "the formal conditions on ρ obtain in cases that aren't genuine reduction pairs. Hence even with the added conditions beyond mere isomorphism... structuralist reduction remains inadequately weak.

[22]Notice how this illustrates the external problem as set out in chapter 1.2. Without a criterion of reduction to hand, how are we to determine that these historical case *are* actual cases of reduction without begging the question?

> Isomorphism is too weak... but so are the stronger set-theoretic conditions structuralists have built into their accounts" (Bickle, 1996, 76)[23]

For Bickle's formal model to be tenable, Bickle needs to solve this problem. I shall now set out how Bickle purports to do this and show why this fails to be a persuasive solution. So, how does Bickle propose to solve this problem? The details are, unfortunately, rather involved:

> "Construe the global reduction relation ρ, on sets of potential models, as constructed out of the "local" links between the elements making up the ρ-related potential models of the respective theory elements. That is, construe ρ as an *ontological reductive link* (ORL)." (Bickle, 1996, 77)

To understand ORLs fully we need to 'unpack' potential models further. $x \in M_p(T)$ is an ordered (n+m+p)-tuple of the form

$$x =< D_1, ..., D_n, A_1, ..., A_m, ...r_1, ..., r_p > \tag{2.1}$$

where:

> "the D_i are the "real" or "empirical" base sets, the A_i are the "auxiliary" base sets (mathematical or other formal spaces), and the r_i are are the theory's fundamental relations *typified* by the base sets." (*ibid.*)

Then ρ is an ORL if it

> "consists of relations between each of the empirical base sets of [to-be-reduced theory] T_O and some element or elements of the potential models of reducing [theory] T_B... Intuitively, ORLs consist of links between the empirical base sets of the [to-be-reduced] theory's potential models and ρ-related elements of the reducing theory's potential models. *The intended empirical applications of the respective theories induce these links.*" (*ibid.* emph. added.)

Bickle then proffers a couple of examples:

[23]This is echoed by Balzer et all, too: "mere formal comparison cannot determine any kind of reduction." (Balzer et al., 1987, 264)

> 'The reduction relation ρ must meet the appropriate formal conditions, and ρ *must also be an ORL*. That is, it [ρ] must be constructed out of local links obtaining between all of the empirical base sets of the [to-be-reduced] theory and elements of the potential models of the reducing theory in a way that respects how the two theories each carve up the world... [In the case of exchange economics and thermodynamics,] the ontologies of the two theories... enjoy no links to one another [whereas] the set of phlogiston elements... gets related via an ORL to the empirical base set of chemical elements containing oxygen... [The] local links comprising the global relation ρ in genuine reductions meeting the ORL condition are not a matter of mere contrivance." (Bickle, 1996, 81)

I contend that is just a piece of set-theoretic sophistry! Formally, ORLs are mappings between the base sets of the potential models of each of theories. But, of course, not any mapping will do: the whole point of requiring ρ to be an ORL is to 1) avert the possibility of the reduction relation holding between totally unrelated theories (to avoid the spurious reduction problem) and 2) to ensure that the 'right' mappings occur between elements of the respective base sets. But just how "the intended empirical applications of the respective theories *induce* these links" (op. cit.) is never explained. It is merely stipulated, for example, that in the "exchange economics to thermodynamics case" that they "enjoy no links to one another" (op. cit.) The formalism here is simply not doing any work. Quite the opposite in fact: it is obfuscating exactly the substantive issue. *At best*, all that Bicke has shown is that once we know how to 'connect' the two theories together - that is, once we have the relevant bridge-laws to hand - ORLs can be used to assimilate this into the structuralist machinery. Of course once this is seen, it pulls the rug from under the much lauded claims about NWR, namely that there is no recourse to bridge-laws. And it really is the case that this claim is repeated time and again. Here is just one exemplyfying passage:

> "Since structuralists construe theories in terms of their models instead of their linguistic expressions, analogs of Nagel's "derivability" and "connectability" conditions nowhere require problematic analogs of [bridge-laws]." (Bickle, 1996, 65)

As with the informal accounts, so with the formal one, this mantra simply

does not ring true. To take stock of the formal account: To reduce one theory to another we first need to give structural renditions of each theory. We then characterize T_O^* which is a subset of the potential models of the reducing theory. We have a reduction of T_O (the to-be-reduced theory) to T_B (the reducing theory) if an analog relation, AR, holds between T_O^* and T_O. (This is what Bickle repeatedly lauds as Hooker's insight.) AR in structuralist terminology is called ρ and is mapping between the elements of the potential models of each of the theories. This mapping needs to meet certain formal criteria and be an ORL. It needs to be an ORL to avoid spurious reductions. But other than the functional characterisation given for an ORL - namely that it functions to connect reductive theories in the right way - it remains entirely mysterious. We are told that "the intended empirical applications of the respective theories *induce* these links" (op. cit.) but just how they do so is never explained.

It is clear that, just as with the informal models, the structuralist turn in its formal guise has simply obfuscated the important components involved in a reduction and certainly does not avoid the purported problems associated with the Nagelian model for those problems find their analogs in the formal model.

Finally, let us consider what Bickle promoted as the upshot of the formal new wave model over the informal ones, namely allowing for a precise characterisation of where a reduction falls on the intertheoretic-reduction spectrum. How is the relative smoothness of a reduction to be captured formally? It is captured in terms of 'blurs'.

A blur is an element of a uniformity (or uniform structure), where uniformities are sets of sets of ordered pairs of potential models. I wish to avoid getting bogged down in needless set-theoretic jargon, but roughly speaking a blur is an element-wise change on a set, such that these changes make one theory identical to another. For example, imagine you have two sets of elements. Think of an element-wise change as replacing one element from the first set with a new element. A blur is a list of element-wise changes needed to make the first set identical with the second. We can then speak of the 'size' of blurs for pairs of sets. The crux of the issue is then this: why is the size of a blur between the reducing and to-be-reduced theories (once they themselves have been suitably formalised) a suitable proxy for where on the inter-theoretic spectrum that putative reduction falls? Again this is surely just set-theoretic sophistry! The 'size' of a blur does not give any sense of the counterfactualness of the the auxiliary assumptions at

all - it simply tells you which elements in one set to replace with another so as to make the first set identical with the second. But how this is a measure of the counterfactualness of the auxiliary assumptions is entirely opaque.

In any case, even the previous point notwithstanding, the structuralist would need to provide a criterion for the admissibility of the size of blurs. Without such a criterion every pair of (formalisable) theories would be on the inter-theoretic reduction spectrum because with enough element-wise changes you can get from one theory to any other! But, as Bickle himself notes, "other context-dependent and mostly pragmatic considerations figure into determining admissible blurs." (Bickle, 1996, 86). So even in this respect, the formalised NWR provides no better a model of reduction.

2.4 Chapter Summary

Advocates of New Wave models of reduction claim that such models are the right alternative to the Nagelian one. I have shown that they are not. In both its informal and formal guises, NWR does not avoid the *internal* problems besetting Nagelian reduction, and, in fact, obfuscates the key issues that are at stake. If it is not a tenable alternative to Nagelian reduction, then it is not a competitor to the Neo-Nagelian model.

Chapter 3

Multiple Realizability

3.1 Chapter 3 Introduction

Multiple realizability is an issue which shadows all discussions of reduction. It is a source of great confusion and controversy, and it is an understatement to say that there is not a consensus about it. Indeed, 'it' is a misnomer, for there are many different positions with regards to multiple realizability. In order to try to give a sketch of the overall landscape, I'll write generally of arguments 'from' multiple realizability to denote those arguments, which involve multiple realizability and that putatively impinge on reduction.

The most well-known and cited argument that involves multiple realizability comes from Fodor's "Special Science. Or: The Disunity of Science As A Working Hypothesis" (Fodor, 1974). This paper has been the touchstone for virtually every paper about reduction and reductionism ever since. Fodor's arguments therein can be seen as a development of an argument presented by Putnam (1967). But whilst Putnam's was specifically concerned with multiple realizability in the context of philosophy of mind, Fodor was concerned with multiple realizability across the 'special sciences', i.e. all the sciences excluding physics.

Fodor's is an argument against reductionism, the thesis that the 'special sciences' reduce to physics. He is of the view that reductionism is ultimately an empirical thesis, yet given what we do know, Fodor argues, reductionism is likely to be false. It is important to be clear from the outset that Fodor's argument is not one which purports to show that a *particular* theory does not reduce to another. It is a general argument in that it threatens to undermine every putative instance of reduction. In particular, it threatens to undermine every putative

reduction in physics too.

To get a sense of the lay of the land, it is useful to give a very rough sketch of Fodor's argument: if there is a multiply realized property in a 'higher-level' theory then this is likely to block the putative reduction of it a 'lower-level' theory. It is very likely that there are such properties. Thus, it is very likely that that theory will not reduce to the 'lower-level' theory and, by extension, reductionism is likely to be false. (Throughout, Fodor's talk of 'higher-level' and 'lower-level' theories should be taken as the putative to-be-reduced and reducing theories. Likewise 'higher-level' and 'lower-level' properties, the properties of the to-be-reduced and reducing theories respectively.)

Clearly a lot more needs to be said about all the notions involved in Fodor's argument but with this coarse characterisation to hand, the various opposing positions can be cast. Contra Fodor, Lewis (1969), Sober (1999), and Richardson (2008) have argued that multiple realizability is not incompatible with reductionism i.e. that a multiply realizable property would not block the putative reduction. More specifically, Lewis argues that particular cases of multiple realizability are compatible with a contextually sensitive or 'local' reduction and that therefore a 'local' kind of reductionism is defensible despite multiple realizability. Richardson argues that Fodor is misrepresenting the Nagelian model of reduction that underpins the entire discussion: by Fodor's lights, *given* the Nagelian model of reduction, multiply realizable properties entail that the to-be-reduced theory in which they appear cannot be reduced to the reducing theory. Richardson argues that the Nagelian model does not entail this and that therefore MR and reductionism are compatible. Sober engages with the notion of explanation, as it features in Fodor's argument. Sober's reading of Fodor's argument is that multiple realizability undermines explanation. The 'special sciences', in virtue of having multiple realizable properties, have an explanatory power that cannot be captured by the 'lower-level' science, and it is this that preserves their autonomy and undermines reductionism. Sober argues that this is not the case and that therefore reductionism and multiple realizability are compatible.

Other authors have argued that Fodor's argument poses a problem for the Nagelian model of reduction, and that it prompts a new model to be put forward. The thought is that reduction in the Nagelian sense is vulnerable to the multiple realizability argument but that just those instances which are undermined *are* genuine cases of reduction, intuitively speaking, so a new model of reduction is

needed to do justice to intuition. Kim (2000) has argued for a new model, which he calls 'functional reduction'. Bickle (1996) has argued that the New Wave model of reduction avoids the problems stemming from multiple realizability.

A different kind of response to MR is to focus on the notion of 'multiple realizability' itself. Indeed several authors argue that multiple realizability is a vague notion and that it can be explicated in a number of different ways. Polger (2008) argues that there are four different intuitions that underpin MR and Lyre (2009) argues, albeit metaphorically, that MR is itself multiply realized! Depending on the particular notion, reductionism is or is not undermined, it is claimed.

Related to the question of what MR really *is*, is the question of how much of it there is about. There are authors who think that (the various kinds of) MR is (are) ubiquitous in science. This is taken as indicative of the falsity of anti-reductionist arguments from MR for, so the counter arguments go, science is, as a matter of fact, full of reductions. Bickle (1996), Clapp (2001) and Endicott (2005) have arguments along these lines. A particular version of this argumentative strategy is that of Enc (1983), who argues that the best case of reduction, viz. thermodynamics to statistical mechanics, is one that involves a multiply realizable property, temperature. This, Enc claims, shows that MR cannot be a stumbling block to reduction. However, Lyre (2009) argues that the case of temperature is the best example (as opposed to mental properties, which are the standard example) of why multiple realizability *is* a problem for reduction!

Shapiro (2000) goes against this grain: he argues that there are good reasons to think that, far from being ubiquitous, the various putative instances of multiply realized properties are not in fact multiply realized after all. He does so by proffering a criterion for 'genuine' multiple realizability, which turns on how lower-level properties realize their higher-order correlates. Given this criterion, Shapiro argues that there are no genuine cases of multiple realizability and that therefore, *de facto*, reductionism is not blocked. A narrower, but similar argument, is made by Bechtel and Mundale (1999). The most cited putative instances of MR properties come from psychology/philosophy of mind. For example, pain is often cited as a property that is multiply realizable and it is this example that both Putnam and Fodor use to motivate their positions. However, Bechtel and Mundale (1999) argue that neuroscience, in fact, supports the claim that pain is not a multiply realizable property.

This is just a small sample of the plethora of different positions pertaining to the issue of MR, but they cover the key positions in the literature. In this thesis I am not concerned with the truth or falsity of reductionism. Whether or not everything reduces to physics remains an open question, but I set this question aside. However, as I have argued in chapter 1, on the Neo-Nagelian account whether one theory reduces to another is not an empirical matter, but a theoretical one. By extension, reductionism is not an empirical thesis but a theoretical one. This is reaffirmed below.

In this chapter I will consider what problems, if any, there being multiply realized properties in a 'higher-level' (to-be-reduced) theory poses for its reduction to a 'lower-level' (reducing) theory.

I start out by considering Fodor's argument against reductionism in section 3.2. I argue that it has no purchase against Neo-Nagelian reduction (and indeed against Nagelian reduction, which was its intended target). This will set the stage for the broadening of the discussion. In section 3.3, I consider the claim that there being multiply realized properties in the to-be-reduced theory undermines the explanatory import that a Neo-Nagelian reduction affords. I argue that this claim is erroneous. In section 3.4 I consider the issue of ontological simplification. It is a widely-held position that there being multiply realized properties in the to-be-reduced theory means that a reduction of that theory to the reducing theory prohibits ontological simplification. I shall consider this claim against Nagelian reduction. In chapter 1.4.7, I argued that the usual way in which Nagelian reduction - *Simplistic Ontological Simplification* (SOS) - is taken to provide ontological simplification is simplistic and metaphysically unattractive. Following on from chapter 1, I show that the way Neo-Nagelian reduction affords ontological simplification is unaffected by multiple realizability.

Before turning to Fodor's argument against reductionism, it is pertinent to specify what it is for a property to be multiply realized. Here is the broadest characterisation:

> *Multiple Realizability*: One type of property is multiply realized iff it is realized by at least two different types of properties.

Three remarks about this characterisation are due: First, notice that multiple realizability pertains to property types, not tokens. Second, 'different' is for emphasis; it is strictly speaking superfluous. Third, and most importantly,

'realization' is to be taken as a place-holder for the relation between the properties. Just what the relation between the properties is is what is at stake. Indeed, Fodor's multiple realizability schema (below) is intended to show that under *different* readings of the 'realization' relation, i.e. identity, nomic connection or correlation, the reduction of one theory to another fails.[1]

Throughout the rest of the chapter I will use 'MR property' to denote a property which is (at least, putatively) multiply realized, and 'MR bridge-law' for a bridge-law which takes as its arguments a MR property and a disjunction of its realizers, respectively. Other obvious abbreviations will also be used: 'non-MR property' and 'non-MR bridge-law'.

3.2 Fodor's Argument Against Reductionism

Fodor's 'Special Sciences' paper (Fodor 1974) is the touchstone for any discussion of reduction and multiple realizability. Fodor's paper is extremely intricate and idiosyncratic in its nomenclature. I will avoid a hermeneutic approach to it and instead I present the core argument against reductionism.

Fodor gives an argument against reductionism 'from' multiple realizability.[2] Whilst Fodor claims that reductionism is 'ultimately' an empirical thesis, the argument he presents putatively shows it likely to be false. More specifically, there being multiply realized properties in the to-be-reduced ('higher-level', in Fodor's terminology) theory makes it unlikely that it reduces to the reducing ('lower-level') one.

To able to assess Fodor's anti-reductionist claims one needs to consider the following question: What notion of 'reduction' underpins Fodor's paper? It is only by settling on a notion of reduction first, that one can assess whether or not reductionism is likely to be false, as Fodor claims it is. It is striking that Fodor does not cite *any* models of reduction in this paper. (In fact, there are only two bibliographical entries in the original paper, one being Block and Fodor (1972) and the other being Chomsky (1965).) However, as will become clear below, the

[1] Sometimes *realization* is taken to be distinct kind of relation between two properties, and not a place-holder. See, for example, Gillett (2003). This seems misguided to me.

[2] Somewhat oddly Fodor seems to use 'reductionism' and 'reductivism' interchangeably. I'll use 'reductionism' from here on as this is how most of the literature since has dubbed the thesis. Nor does Fodor talk explicitly of multiple realizability rather 'instantiations' of the higher-level properties by the lower-level ones. But, again, this is just a matter of terminology and using this term makes Fodor's paper congruent with the modern nomenclature.

model of reduction underpinning his paper is the Nagelian one: reduction consists in the derivation of the law of the 'higher-level' from the 'lower-level' theory, and this derivation involves bridge-laws. Both are hallmarks of Nagelian reduction. Indeed, there is a broad consensus that Fodor is presupposing the Nagelian model. cf. for example, Gillett (2003), Polger (2008), and Sober (1999). My concern in this thesis is not to defend Nagelian reduction, of course - I am advocating an alternative, Neo-Nagelian reduction. However, the arguments which Fodor presents are equally *applicable* to the Neo-Nagelian model, for they pertain to just those features which the Nagelian and Neo-Nagelian models have in common. I shall argue that Fodor's argument is ineffective against both Nagelian and Neo-Nagelian reduction.

3.2.1 Stage-Setting

Here is Fodor's schema for reduction. Consider laws of a 'special' or 'higher-level' theory to be of the form:

(1) $S_1x \rightarrow S_2x$

This is "intended to be read as something like 'all S_1 situations bring about S_2 situations'." (Fodor, 1974, 98) Fodor notes that the 'all' 'needs to be taken with a pinch of salt' (*ibid.*) because the 'higher-level' laws are not exceptionless. (I return to this below.) He then proffers the following necessary and sufficient conditions for (1) to reduce to a law of physics, namely that there exist the following three laws:[3]

(2a) $S_ax \rightleftarrows P_ax$

(2b) $S_bx \rightleftarrows P_bx$

(3) $P_ax \rightarrow P_bx$

where the P_s stand for predicates of physics, (3) is a physical law and (2a) and (2b) are 'bridge-laws'. Fodor notes that the characteristic feature of (2a) and (2b) is that they contain predicates from both theories. Here it is supposed that the higher-level properties are not multiply realized and, as such, (2a) and (2b) are non-MR bridge laws.

[3]I have slightly modified Fodor's sub- and super-scripts for the sake of clarity.

Now consider the case where the higher-level is multiply realized. In this case, the bridge-laws take the form:.

(4a) $S_a \rightleftarrows P_a^1 \vee P_a^2 \vee \ldots P_a^n$

(4b) $S_b \rightleftarrows P_b^1 \vee P_b^2 \vee \ldots P_b^n$

as well as a set of corresponding physical laws:

(5) $\{P_a^i \rightarrow P_b^i\}$ (for $i = 1, 2, \ldots n$)

Although Fodor has characterised how the connectives (arrows) in (1) and (3) are to be read, as above, there are "quite serious open questions about the interpretations of [the connective] in bridge-laws" (Fodor, 1974, 99). Fodor rejects the possibility of reading the connective in bridge-laws as 'brings about' or 'causes' for these are asymmetric "while bridge laws express symmetric relations."(*ibid.*) Syntactically speaking, then, bridge-laws cannot be conditionals; Fodor is committed to their being bi-conditionals, although just what they express is a further question to be answered.

As I said at the outset, Fodor's argument is intricate (often tortuous) and riddled with an idiosyncratic nomenclature, and there are several arguments that bleed into one another so some clarification is needed. I identify three distinct arguments. Two are what might be called, 'direct' arguments against reductionism. The first of these is an implausibly strong one which is easily rebutted. I do so in section 3.2.2. The second is Fodor main argument against reduction, and I deal with this in section 3.2.3. Fodor also argues that MR-properties prevent ontological simplification, but I shall consider this argument separately in section 3.4.

3.2.2 The 'Modal' Argument Against Reductionism

Fodor has a 'modal' argument against reductionism, which turns on the interpretation of the connectives in the bridge-laws.

> "[If the connective in the bridge-law] is interpreted as any relation other than identity, the truth of [reductionism] will only guaranty the truth of a weak version of physicalism, and this would fail to express the underlying ontological bias of the reductivist program." (Fodor, 1974, 99)

Why does a relation other than identity only guarantee a weak form of physicalism? Fodor's argument is that short of identities:

> "[bridge-laws] claim at most that, by law, x's satisfaction of a P predicate and x's satisfaction of an S predicate are causally correlated [but] this is compatible with a non-physicalist ontology since it is compatible with the possibility that x's satisfying S should not itself *be* a physical event." (Fodor, 1974, 99)

Fodor is going to go on to argue that multiple realizability shows that bridge-laws cannot be identities and hence conclude that reductionism is likely to be false. Obviously, this argument is close to the argument against ontological simplification (cf. section 3.4) but I wish to prise them apart for the sake of clarity. The point I want to focus on here is the modal one. Fodor thinks that bridge-laws expressing anything other than identities do not rule out the *possibility* of a dualism vis-á-vis the ontology of the special science and that of physics. By Fodor's lights bridge-laws which 'fall short' of identities are too weak; too weak because they do not *guarantee* ontological simplification. As I said, whether bridge-laws-cum-identities do provide ontological simplification or whether a lack of such bridge-laws precludes it is something I return to below. But here let us just focus on the modal claim.

I think that the right reaction to this argument is incredulity. Surely tying the truth or falsity of reductionism to the preclusion of dualism is, too stringent a requirement. Notice that this is far stronger than requiring that a successful reduction *actually* afford ontological simplification: Fodor is construing reductionism as requiring dualism to be shown to be impossible. (It is worth emphasising that the possible lack of identities *does not imply* dualism; rather the lack of identities *does not preclude the possibility* of it.) What precedence is there for such a strong construal of reductionism? I cannot find any, and it strikes me as intuitively uncompelling.

Intuition aside, the problem with this argument is that it is not consistent with the position that Fodor is advocating, namely that reductionism is ultimately an empirical thesis. For on this reading, reductionism is not an empirical thesis but, so to speak, a modal one too! For these reasons, I shall not consider *this* argument against reductionism any further. Reductionism does not rule out the *possibility* of dualism, and it is a better thesis for that, I suggest.

3.2.3 The Main Argument Against Reductionism

Having dealt with the 'modal' argument, I will now go on to consider what I take to be Fodor's main argument against reductionism. I present what I think is a fair reconstruction of Fodor's argument, which omits certain unnecessary parts of Fodor's paper. The overall structure of Fodor's argument is as follows:

> (MR1) It is likely that the laws of the 'higher-level' science do involve multiply realized properties.
>
> (MR2) The 'higher-level' ('special') sciences are unlikely to reduce to physics, if the laws of the former involve multiply realized properties.
>
> (MRC) The 'higher-level' science is unlikely to reduce to physics.

The argument is valid[4]; the question is whether it is sound. Notice that there are two recourses to the notion of likelihood in the premises: first, it is *likely* that the properties of the 'higher-level' theory are multiply realized (MR1); second, if so, the 'higher-level' theory is *unlikely* to reduce (MR2). Each of these underpins one sense in which reductionism is taken to be an empirical thesis, as follows.

3.2.3.1 MR1

It seems to be a straightforwardly empirical question whether or not a higher-level property is multiply realized and, Fodor contends, it is likely that there are MR properties in the 'higher-level' sciences. For the laws are:

L1 about tokens of properties[5] 'whose physical descriptions have nothing in common' (Fodor, 1974, 102);

L2 and, often, whether or not the physical descriptions subsumed under the generalizations have anything in common is 'in an obvious sense, entirely irrelevant to the truth of the generalizations, or to their interestingness, or to their degree of confirmation or, indeed, to any of their epistemologically important properties' (*ibid.*);

[4]Well, at least if 'likely' is dropped from MR1. More on this caveat shortly.

[5]Fodor idiosyncratically uses 'events' to denote tokens of properties and 'properties' as the kind under which the tokens fall. The given notation is clearer.

He takes it that "these remarks are obvious to the point of self-certification" (*ibid.*) *Pace* Fodor, I do *not* think that whether or not a property is multiply realized *is* an empirical question for bridge-laws (MR, or otherwise). This will become clearer when we consider the question of the ontological simplification in section 3.4 below. But suppose for the sake of argument that whether or not there are multiply realized properties is an empirical issue. Is it the case that there are likely to be such properties in the to-be-reduced theory? A cursory consideration of L1 and L2 shows them *not* to support the likely truth of there being such properties.

L1 simply begs the question: that the tokens of the properties whose *physical descriptions which have nothing in common* is no more than an elliptical expression for saying that tokens of the 'higher-level' property are realized by tokens of different physical kinds. That is, L1 just states that the 'higher-level' property is multiply realized and so cannot be an argument for the likelihood of there being MR properties.

L2 also does not support the claim that MR properties are likely. Even if we grant what Fodor says here – and this is to grant a lot: is it 'obvious', for example, that it is entirely irrelevant to the truth of the 'higher-level' generalization (i.e. special science law) whether or not the properties that it is couched in are multiply realized? – none of this speaks to the likely truth of there being multiply realized properties in the 'higher-level' theory. The caveat in L2, which I should say is a paraphrase of Fodor's own (cf. Fodor (1974, 103)), attests to this: the "epistemologically important properties" of the generalisations are impervious to *whether or not* the properties in terms of which they are couched are multiply realized or not. So considerations of the "epistemologically important properties" cannot speak to the likely truth of multiple realizability.

Whilst L1 and L2 might not speak to the likelihood of multiple realizability, it may be argued that it is, nonetheless, likely that multiply realized properties abound. After all, isn't it just obvious, intuitively speaking, that the properties that Fodor cites, such as pain and monetary exchange (discussed below), are multiply realized? Certainly, pain seems to be realized in *distinct physical systems*, say in humans, dogs, molluscs and so forth. But the devil is in the details here and the details are always contentious. For one thing, the physical systems may not be distinct in the relevant sense. For example, pain in both humans and dogs might be realized by c-fibre firing, which is what is often cited as the lower-level

realizer of pain in humans, and thus pain may not be multiple realized after all, it could be argued. (This line of argument is pursued by Shapiro (2000).)

For another, any putative case of a multiply realized property could be construed as one in which the putative singular 'higher-level' property is in fact distinct. For example, it might be argued that if human pain and dog pain are in fact realized by distinct neurological kinds (say, c-fibre firing and d-fibre firing) then this is good reason to take human pain and dog pain to be distinct higher-level properties. (Kim runs some such argument - Kim's well-know 'Jade' example. cf. section 3.4.2.)

In reply, it would presumably be argued that pain is taken to be a singular property precisely because the laws about human pain, dog pain, and mollusc pain are the same. This is an important point, not only for the present issue but the discussion at large. Putative higher-level properties are taken to be a single property precisely because they feature in some law. But the matter is not straightforward: first, the laws about pain (be they about humans, dogs, or molluscs) are not exceptionless. Second, not every law about, say, human pain holds true of the others, either in degree or in kind. For example, suppose that there is a law that x amount of pain brings about a quickening of heart-rate by y in humans. Maybe there is a correlation in dogs too, but it strikes me as unlikely that it would be exactly the same (i.e. the values of x and y would be different.) And this law does not even apply to molluscs, which have no hearts. Thus the very thing which underpins the putative singularity of the higher-level property is called into doubt.

More importantly, the questions about pain to which I am alluding are not simply empirical questions either. The extra-empirical matter of establishing whether an animal is in pain, irrespectively of the consequent of the conditional that forms the law that one wants to verify, say, is controversial. All of this is merely suggestive, and I am not putting forward the position that there are not any interesting generalizations about 'pain', which, depending on ones' view about laws, may be laws. I am merely pointing out that the intuition that multiply realized properties abound is a tricky one to articulate, let alone defend. Intuitions seem to cut both ways; the literature is filled with conflicting examples.

My view is that the starting assumption that multiple realizability is an empirical matter, i.e. the assumption that whether or not a particular property is multiply realized is a matter of fact, is erroneous. And this is for the same rea-

son that bridge-laws in general are not factual claims, or as I put it elsewhere, bridge-laws are not metaphysically substantial. Once this is noticed, many of the conflicting intuitions about whether or not a particular property is multiply realized abate, or so I shall argue.

For the sake of the current argument though, let us suppose that the properties in the higher-level science are multiply realized and turn to the second premise, MR2.

3.2.3.2 MR2

MR2 says that if there are multiply realized 'higher-level' properties, then it is unlikely that the 'higher-level' theory will reduce. What Fodor has in mind is best seen by considering his example of Gresham's law. It is instructive to quote the relevant passage in full.

> "Suppose, for example, that Gresham's 'law' really is true. Gresham's law says something about what will happen in monetary exchanges under certain conditions. I am willing to believe that physics is general *in the sense that it implies that any event which consists of a monetary exchange* (hence any event which falls under Gresham's law) *has a true description in the vocabulary of physics and in virtue of which it falls under the laws of physics.* But banal considerations suggest that a description which covers all such events must be wildly disjunctive. Some monetary exchanges involve strings of wampum. Some involve dollar bills. And some involve signing one's name to a check.... What are the chances that a disjunction of physical predicates which covers all these events expresses a physical natural kind? In particular, what are the chances that such a predicate forms the antecedent or consequent of some proper law of physics? [...] A natural kind like a monetary exchange could turn out to be co-extensive with a physical natural kind; but if it did, that would be an accident on a cosmic scale." (Fodor, 1974, 103-104)

Note that we are to suppose that Gresham's law really is a *proper* law.[6] According to Fodor, Gresham's law does have a true physical description, and,

[6]Just what is for something to be a proper law I return to shortly, but the contrast Fodor has in mind is, I take it, that of an accidental generalization. In any case, under any sensible construal of that term, this is not an obviously innocuous supposition. That is, I think there

thus, Gresham's law 'falls under' physical laws. However, the physical description which covers each instance of Gresham's law ('all such events') 'must be wildly disjunctive' for the physical description will differ from those instances of Gresham's law that involve dollar bills, to those involving checks, and wampum, and so forth. One can describe all the instantiations of the law in terms of a disjunction of physical predicates but, and this is the crucial point, it is unlikely that this disjunction of predicates forms a physical natural kind.

This last claim is crucial. To understand it, it is necessary to consider Fodor's characterisation of natural kinds. Fodor proffers the following:

> "[T]he natural kind predicates of a science are the ones whose terms are the bound variables in its proper laws. I am inclined to say this even in my present state of ignorance, accepting the consequence that it makes the murky notion of a natural kind viciously dependent on the equally murky notions *law* and *theory*." (Fodor, 1974, 102)

This obviously prompts the question of what a theory's proper laws are. Fodor writes:

> "[A] necessary condition on a universal generalization being lawlike is that the predicates which constitute its antecedent and consequent should pick out natural kinds." (Fodor, 1974, 108) [7]

By Fodor's lights, then, monetary exchange is a natural kind because it appears in a *proper* law viz. Gresham's law. (Recall that Gresham's law is a 'proper' law *by supposition.*) Whilst it is possible (or better: 'not demonstrably *impossible*') that this kind is co-extensive with a physical natural kind, it is incredibly unlikely. Why? For it be, the entire disjunction of physical predicates that describes all instantiations of Gresham's law would have to appear, by the same criterion, in one of the 'proper' laws of physics.

Actually, the situation is worse for the reductionist, argues Fodor, for:

> "[reductionism] claims not only that all natural kinds [are, if not identical with, then at least] co-extensive with physical natural kinds, but that the co-extensions are nomologically necessary: bridge laws are

are good grounds to think that Gresham's Law is not a law of nature. But let us grant this for the moment.

[7] A bizarre consequence of this is that a bridge-law *qua* identity counts as proper law!

> *laws*. So, if Gresham's law is true, it follows that there is [i.e. there must be] a (bridge) law of nature such that 'x is a monetary exchange $\leftrightarrows$ x is P', where P is a term for a physical natural kind. But, surely, there is no such law." (*ibid.*)

All this is not particular to economics of course. Fodor argues for the same conclusion in the context of psychology:

> "Even if (token) psychological events are (token) neurological events, it does not follow that the natural kind predicates of psychology are co-extensive with the natural kind predicates of any other discipline (including physics)... [It is doubtful that] there are neurological natural kinds co-extensive with psychological natural kinds. [And moreover,] even if there is such a co-extension, it cannot be lawlike [sic.]. For, it seems increasingly likely that there are nomologically possible systems other than organisms (namely, automata) which satisfy natural kind predicates in psychology, and which satisfy no neurological predicates at all." (Fodor, 1974, 105)

This last passage betrays a minor inconsistency in Fodor's exposition: 'events', as Fodor has used the term, are tokens of 'properties'. So a psychological event is a token of a psychological property and a physical event is a token of a physical property. So, speaking of 'token neurological events' is a category mistake, for the events do not themselves form types, of which there are tokens - the events are the tokens of types of properties. But *mutatis mutandis*, by Fodor's lights, even if one can give a description of psychological events in terms of neurological events, it does not follow that the natural kinds of psychology are even co-extensive (let alone, identical with) the natural kinds of neurology. Why? The description of the psychological events in terms of the neurological events will be 'wildly disjunctive' and this disjunctive predicate is unlikely to appear in any of the laws of neurology; it is unlikely that it will be a natural kind of neurology. It is therefore unlikely that the psychological natural kind is co-extensive with a natural kind of neurology. Moreover, even if by an accident on a cosmic scale it were the case that it is co-extensive with a neurological natural kind, this co-extension, Fodor would implore, is *surely* not law-like for there are possible systems for which psychological laws would hold true but which are entirely neurologically (micro-physically) distinct.

To summarise Fodor's argument: whenever the 'higher-level' law involves multiply realized properties, the right-hand-side of the requisite bridge-laws, as per 4a and 4b, is unlikely to be a physical natural kind term. In all likelihood then, ontological simplification will not be had, for, *ex hypothesis*, the left-hand-side of the bridge-law *is* a ('higher-level') natural kind term and hence cannot be identical with the former. For the same reason, viz. the disjunction on the right-hand-side, the requisite bridge-laws are unlikely to be proper laws, for in order to be a proper law its arguments need to be natural kind terms! Fodor takes ontological simplification and bridge-laws being 'proper' laws to be necessary conditions for reduction. The punch-line of all this is now clear: multiple realizability is likely to prevent a putative reduction because ontological simplification is undermined as the bridge-law cannot be taken to express identities and the requisite bridge-laws will fail to be proper-laws.

3.2.4 Fodor's Argument Repudiated

Fodor's main argument against reductionism is unpersuasive for three reasons. First, it turns on a problematic definition of natural kind terms and laws. In section 3.2.4.1, I show that the definitions proffered show Fodor's argument to be *too weak* to establish his desired conclusion. Second, Fodor's position is not internally consistent: given the aforementioned definitions it is not the case that reductionism is *likely* to be false, it is false by definition! This is contrary to the premise that reductionism is an empirical thesis. This is in section 3.2.4.2. Most important given present concerns, in section 3.2.4.3, I show that Fodor's argument misses its target: Nagelian reduction simply does not require that bridge-laws be 'proper' laws with any sense of propriety. For the same reason, Fodor's argument does not impinge upon Neo-Nagelian reduction.

3.2.4.1 Too Weak An Argument

Reconsider Fodor's definition of a natural kind term: a natural kind term is a term that appears in a law of nature, a 'proper' law.[8] Fodor by way of an admission remarks that this ties the 'murky' notion of 'natural kinds' to the 'equally murky' notion of 'law' (*op. cit.*). But the problem here is not that the latter notion is

[8]Fodor actually relativizes natural kind terms *to* theories in that he speaks of a particular theory's natural kind terms. I take it that this is a *façon de parler* and not that Fodor is committed to the theory-relativity of natural kind terms.

murky *per se*. The problem is that the very notion of 'law' is in turn defined in terms of natural kinds: a necessary condition for a true generalization to be a law, we are told, is that its arguments be natural kinds. This is obviously problematic from an epistemological point of view: there is no independent way to check whether the bridge-laws are proper laws. But granting this characterisation of the terms, one can show that Fodor's argument is *too weak* to establish his desired conclusions.

Fodor's characterisation of the relevant terms is *too weak* because it *fails* to show that bridge-laws are not proper laws in case of multiple realizability, as Fodor's purports to show. One is told that in case of multiple realizability the requisite bridge-law is likely not to be a proper law because the disjunction on the right-hand-side is likely not a natural kind term. Yet, that the disjunction is likely not a natural kind can only mean that it is not likely to be an argument in a 'proper' law. So from this one must conclude that the bridge-law is not a 'proper' law. But this fails to show that bridge-laws are not proper laws: this can only mean that it is not a 'proper' law because it is not a 'proper' law! The inter-definition of the relevant terms robs Fodor's argument of any purchase. Fodor's reply would, presumably, be that this gets the order of the argument wrong: the disjunction on the right-hand-side of the bridge law is deemed not to be natural kind term *first*, so to speak, because it does not appear as an argument of a proper law of physics, and this in turn explains why the bridge-law is not a proper law. But given his characterisation of the relevant terms this simply does not cut the mustard: the disjunction does appear as an argument in the bridge-law and, without an *independent* criterion for why the bridge-law does not count as a 'proper' law, one cannot conclude that the disjunction is not a natural kind term.

Fodor might counter that the bridge-law is not a proper law of physics because the argument on the left-hand-side is a 'higher-level' property. To be a 'proper' law of physics the bridge-law must take only physical kinds as its arguments, he might suggest. But if that is correct then no bridge-law is ever going to be a proper law because by construction bridge-laws always take arguments from different theories. The crucial point is that multiple realizability does no work here: this is *irrespective* of multiple realizability. Thus this is no longer an argument that multiple realizability undermines a putative reduction, but that the very use of bridge-laws itself (i.e. including non-MR bridge-laws) undermines reduction.

Thus, Fodor's characterisation of the terms is too weak to establish his desired conclusion, viz. that bridge-laws are not proper laws. I also do not think that bridge-laws are proper laws, of physics or otherwise. As argued in chapter 1.4.7 and as I return to below, bridge-laws are not laws in any proper sense. On the Neo-Nagelian account, bridge-laws are a kind of theoretical stipulation, namely *coherence constraints*. But the point stands: *Fodor's characterisation of the relevant terms* is too weak. For under his characterisation of the relevant terms one can always reverse the reasoning, if you will, for the conclusion that the bridge-law is not a 'proper' law, in the way just indicated.

3.2.4.2 Internally Inconsistent

Fodor's argument, is also internally inconsistent. In particular his characterisation of the relevant terms entails that reductionism is false by definition in case of multiple realizability. This is inconsistent with the assumption that reductionism is 'ultimately' an empirical assumption, as Fodor contends. It is remarkably straightforward to see this: bridge-laws need to be 'proper' laws; to be 'proper' laws, the *disjunction* of the 'P'-properties needs to be a natural kind; but *each* of the 'P'-properties is a distinct kind by the very definition of multiple realizability! This renders MR2 into a new form:

> (MR2') The 'higher-level' sciences *do not by definition* reduce to physics, if the laws of the former take as arguments multiply realized properties.

In case of multiple realizability, then, reductionism fails by definition, contra to the aforementioned assumption. Fodor might reply that reductionism still is an empirical thesis because whether or not there are multiply realized properties remains an open question. That is, he may concede MR2' but still hold on to MR1. But as per section 3.2.3.1 the arguments in favour of MR1 are wanting, and recourse to intuition cuts both ways. For someone committed to reductionism being an empirical thesis, MR2' is surely just too strong a premise.

3.2.4.3 Bridge-laws, 'Proper' Laws, and Natural Kind Terms

Just what 'proper' laws of nature are, and the relation between them and natural kind terms, is a tricky subject. There is a vast literature pertaining to both and an attempt to settle the issue would augment the present work beyond what is manageable. Fortunately, we can undercut this entirely: Fodor's argument is

unpersuasive irrespectively of how one characterises laws of nature - 'proper' laws - and natural kind terms. What is really doing the work (damage) here are the following two assumptions of Fodor's:

> FA1: Bridge-laws need to be 'proper' laws - bridge-laws cannot be disjunctive.
>
> FA2: Ontological simplification can only be had by bridge-laws-cum-identities.

Both of these assumptions are wrong both on the Nagelian and Neo-Nagelian accounts. Let's consider FA1 first. This is assumed throughout Fodor's paper, implicit in MR2 (and MR2'). Recall that Fodor's overall argumentative strategy is that reductionism is unlikely to be true because bridge-laws are unlikely to be 'proper' laws. i.e. the correlation between the 'higher-level' and 'lower-level' properties is is not likely to be law-like. Yet, as set out in chapter 1, both on Nagel's original model and the Neo-Nagelian model I am advocating, bridge-laws are not laws of nature; bridge-laws were never advocated as 'proper' laws with any sense of propriety! (Indeed, the Nagelian model is entirely silent about what count as 'proper' laws and natural kinds, which is not surprising given the philosophical stance that Nagel and his contemporaries took, namely logical empiricism.) In fact, as set out there, Nagel considers various options about the status of bridge-laws, viz. identities, semantic statements or even mere conventions. In the Neo-Nagelian model, I argue against any substantive 'metaphysical' reading of bridge-laws. Bridge-laws are a kind of theoretical stipulation: *coherence constraints*. As regards FA2, Nagel did consider that ontological simplification could be had via bridge-laws as identities. Yet ontological simplification (via bridge-law as identities or otherwise) is not a necessary condition for reduction under his model; at most, it is an extra desideratum. (cf. chapter 1.3.1) The Neo-Nagelian model rejects bridge-laws-cum-identities - indeed, it rejects construing bridge-laws as metaphysically substantial in any sense. A successful Neo-Nagelian reduction can afford ontological simplification, but not via bridge-laws qua identities. In any case, like the Nagelian model, ontological simplification is not a *necessary* condition for Neo-Nagelian reduction. (cf. chapter 1.4.2)

Perhaps pointing out that on both these accounts of reduction, bridge-laws need not be 'proper' laws, and that ontological simplification is not a necessary condition for reduction, will leave the reader cold. After all, isn't Fodor free to define what it takes for one theory to reduce to another? As per section 3.2 there

is good reason to think that Fodor took himself to be engaging with the Nagelian model - the hallmarks of Nagelian reduction are there - rather than proffering an alternative model. There is consensus on this point. However, if a different model of reduction is intended then when Fodor says that multiple realizability undermines the putative reduction of one theory to another, he would not mean 'reduction' as we mean it. All that could be said is that we are talking past one another and given that in this thesis I am primarily concerned with Neo-Nagelian reduction, Fodor's arguments would be largely irrelevant.

In the rest of the chapter, I consider other ways in which multiple realizability putatively impinges on Neo-Nagelian reduction. Specifically, I consider whether multiple realizability undermines the explanatory import that Neo-Nagelian reduction affords. I shall argue that it does not. I will then show that the manner in which Neo-Nagelian reduction can afford ontological simplification is also not undermined by multiply realized 'higher-level' properties. In the final section, I consider some further arguments 'from' multiple realizability which putatively undermine reduction. These too, I shall argue, do not undermine Neo-Nagelian reduction.

3.3 Multiple Realizability and Explanation

Nagelian reduction provides explanation in virtue of the deductive-nomological (DN) model of explanation. Deriving the laws of the to-be-reduced theory from laws of the reducing theory, bridge-laws and auxiliary assumptions just *is* to explain them on the DN account. In chapter 1 I have set out how Neo-Nagelian reduction affords explanation. I provided a broad philosophical framework - pluro-particularism - for explanation and located the explanation that Neo-Nagelian reduction affords as falling under the DN schema. The difference between the Nagelian and the Neo-Nagelian models with respect to explanation is that in the latter there needs to be *warrant* for the bridge-laws and auxiliary assumptions from which the laws of the to-be-reduced theory are derived. So whereas Nagel took bridge-laws and auxiliary assumptions as *givens*, so to speak, the emphasis on the Neo-Nagelian account is to justify the recourse to them.

It is also important to re-emphasise exactly what the explanadum is here: the empirical success of the to-be-reduced theory. To do so we derive its laws for it is the laws of a theory that encode its empirical content. In reducing the theory to

another, we explain why it is that it, the to-be-reduced theory, is as empirically successful as it is (given that it is taken to be strictly speaking false). So it is the empirical success of the to-be-reduced theory, as encoded into its laws, that is the explanadum, and the laws of the reducing theory, the bridge-laws and the various other auxiliary assumptions that are the explanans.

Various authors have argued that multiple realizability poses some problems for explanation for reduction. (cf. Kim (2000), Bickle (1996), Marras (2002), Lyre (2009). How might MR-properties impinge on the explanatory import of reduction? Why is it that MR-bridge laws are considered to be problematic for explanation?

Here are three argumentative strands that one can identify in the literature:

1) MR-properties undermine explanation;

2) The *possibility* of MR-properties undermines explanation;

3) Reduction with MR bridge-laws fails to explain because it lacks 'descriptive richness'.[9]

I shall argue that each of these argumentative strands, once spelled out, is unpersuasive against Neo-Nagelian reduction. They are also ineffectual against Nagelian reduction, for the same reasons. Indeed the arguments here were directed at Nagelian reduction and I am appropriating them against the Neo-Nagelian model, albeit ultimately to reject them.

Before turning to them, I want to stress that in this section I am concerned with whether multiple realizability poses a *special* problem for explanation in the context of reduction. That is, I do not defend the general position that Neo-Nagelian reduction affords explanation via the DN model. (This was developed and defended in chapter 1.4.) Thus here it is supposed that a reduction involving non-MR properties *has* explanatory import, and I consider whether this is undermined if MR-properties are involved instead.

3.3.1 MR Does Not Undermine Explanation

There is a general class of arguments to the effect that MR-properties undermine the explanation that reduction affords. Why? One argument is that MR-

[9]The first is, for example, in Kim (1995), and is criticised, for example, by Marras (2002). The second can be gleaned from Enc (1983). The third is to be found in Fodor (1974) and is critically discussed by Sober (1999), see below.

properties require MR bridge-laws and that the latter are not 'proper' laws, and as such a reduction involving them is unexplanatory. For example, this is how Sober (1999) reads Fodor (1974). But as per the discussion in section 3.2.4.3, we already know that Fodor's argument is ineffectual: that FA1 is not a necessary condition for reduction is, *ipso facto*, why Fodor's argument does not show that MR bridge-laws undermine the explanatory import that a reduction affords. In short, bridge-laws need not be (and indeed are not) 'proper' laws.[10]

A second argument in this broad class is that MR bridge-laws are inadequate as explanans. The point is often made, if not in print (although see, for example, *Bickle1996* and Lyre (2009)), then at least in discussion of the issue; however it is never sufficiently detailed. Why, exactly, is an MR bridge-law not an adequate explanan? Again, the thought is that a disjunctive bridge-law of the form [4a] or [4b] somehow undermines the explanation, but it is not clear why this is so.

Whilst writing specifically with respect to Fodor (1974), I think Sober's sentiment can be generalized when he writes:

> "Are we really prepared to say that the truth and lawfulness of the higher-level generalization is inexplicable, just because [a] derivation is peppered with the word "or"? I confess that I feel my sense of incomprehension and mystery palpably subside when I contemplate [some such] derivation. Where am I going wrong?" (Sober, 1999, 554)[11]

We can sharpen Sober's point: from a logical point of view, the derivation of the 'higher-level' law is valid whether the bridge-laws are disjunctive or not - the derivation 'goes through' irrespectively. Thus, the DN explanation is not adversely affected. It is as simple as that. I cannot find any other (let alone persuasive) arguments to the effect that explanation is undermined simply in virtue of there being MR bridge-laws.

3.3.2 Possible MR Does Not Undermine Explanation

Another argument to the effect that multiple realizability poses a problem for reductive explanation turns, not on actual MR properties, but on their mere

[10] Of course, a DN explanation needs some explanans to be laws but this role is played by the laws of the reducing theory and not the bridge-laws.

[11] I say 'sentiment' because I think that Sober misconstrues the explanadum here. It is not the truth or lawfulness of the higher-level generalization that is to be explained, rather its empirical adequacy. cf. section 3.3.3.

possibility. The argument runs as follows: Higher-level properties are multiply realiz*able* - it is *possible* that a higher-level property is multiply realized. And, even if a higher-level property is known to be multiply realized, it may have further hitherto unknown realizers. But if that's right, then any putative reductive explanation is, at best, incomplete. At best, we have only partially explained the empirical success of the to-be-reduced theory. (This argument does not find explicit articulation in the literature, being usually tied to concerns about ontological simplification. However, this is, for example, an argument that can be gleamed from Enc (1983).)

This is obviously a poor argument, much like Fodor's first argument against reductionism as per section 3.2.2. That an explanation does not explain *every possible* explanandum does not show that it does not explain the *given* explanandum.

3.3.3 Simplicity and Descriptive Richness

A final argument against reductive explanation in case of MR-properties concerns the descriptive richness of the higher-level theory. Sober considers the claim that the higher-level sciences capture patterns "that would be invisible from the point of view of lower-level science." (Sober, 1999, 560) The idea, as Sober puts it, is that the higher-order predicate P captures what all its "instances have in common" in a way that the disjunctive lower-level predicate (A, or A_2 or . .. or A_n) cannot do "in any meaningful sense." (*ibid.*) Sober illustrates the point with the following example:

> "If I ask you what pineapples and prime numbers have in common and you reply that they both fall under the disjunctive predicate 'pineapple or prime number,' your remark is simply a joke." (*ibid.*)

Sober's response is to deny that this point pertains to explanation. In fact, he dismisses it as being irrelevant to explanation:

> Whether or not this claim about the descriptive powers of higher- and lower-level sciences is right, it involves a drastic change in subject. Putnam and Fodor were discussing what higher- and lower-level sciences are able to *explain*. The present argument concerns whether a lower-level science is able to *describe* what higher-level sciences *describe*... However, there is a world of difference between describing a

> fact and explaining the fact so described. This new argument does not touch the reductionist claim that physics can explain everything that higher-level sciences can explain." (Sober, 1999, 560)

I think that this is too quick. First, as stated, Sober's position is underspecified: what sense of 'explanation'? (For the sake of argument, take 'description' to be uncontentious.) Certain kinds of explanations are 'closer' to descriptions than others, it seems to me. Consider mechanistic explanations, which consist in describing the mechanisms that bring about the explanadum. Perhaps this will strike the reader as a perfidious use of 'describe'. I grant that describing the mechanisms is not the same thing as describing the system, for the former requires us to be judicious about what count as the mechanisms in the system. But are we prepared to say that there is 'a world of difference' here? I am not. And furthermore, wouldn't a proximate (complete/total) description of the system be sufficient for answering any explanatory questions about it? It strikes me that it would.

The right response to this problem is not to insist that description and explanation are 'worlds apart', as Sober does. In this context they are not. To see this, and what is really underlying the argument here, reconsider the schema for reduction from section 3.2.[12] Let us suppose that only one higher-level property is multiply realized. i.e. S_a is a MR-property and S_b is a non-MR-property. Then the reduction looks something like the following: we derive (1') from the conjunction of (2a'), (2b') and (3'), where

(1') $S_a \rightarrow S_b$

(2a') $S_a \rightleftarrows P_a^1 \vee P_a^2 \vee \ldots P_a^n$

(2b') $S_b x \rightleftarrows P_b x$

(3') $P_a^1 \vee P_a^2 \vee \ldots P_a^n \rightarrow P_b x$

The point about the descriptive richness is this, I take it. The higher-level theory provides, intuitively speaking, a better explanation of S_b than does the lower-level theory, for the descriptive richness of the higher-level theory provides a

[12] This is, as aforementioned, far too simplistic a schema because it does not include the various auxiliary assumptions. However, it suffices for the present illustration.

simpler explanation of there being S_bs. To make this vivid, compare "$S_a \rightarrow S_b$" to "$P_a^1 \vee P_a^2 \vee \ldots P_a^n \rightarrow S_b$" - an explanation based on the former is surely simpler![13]

There is a shift in discourse here, but it is not the shift that Sober contends it is about. It is about the explanandum. Reductive explanation is not 'empirical' explanation: a reduction does not (not even putatively, that is) explain why certain events occur. A particular law, or theory more generally, may explain why a certain event occurs. Using the example to hand, one might say that the 'higher-level' theory does better explain there being S_bs than the 'lower-level' theory, (i.e. an explanation 'via' $P_a^1 \vee P_a^2 \vee \ldots P_a^n$). But, and this is the crucial point, reductive explanation is not concerned with this: the explanandum in reductive explanation is the empirical success of the to-be-reduced theory, and not whatever further explanatory benefits that that theory affords. It is clear that multiple realizability does not impinge on the explanandum we are here concerned with, viz. the 'high-level' law. In short, the argument 'from' descriptive richness is simply concerned with a different explanandum.

3.4 Multiple Realizability and Ontological Simplification

Perhaps the most persistent argument in this context is that multiply realizable properties in the higher-level science prevent ontological simplification and hence undermine reduction. As I have stressed, however, ontological simplification is not a necessary condition for Nagelian reduction; at most it's a desideratum. Thus even if, broadly speaking, multiple realizability prevents ontological simplification, this does not threaten Nagelian reduction *per se*. Unfortunately, this point is widely unappreciated. Here are two representative quotes. (Also, cf. Bickle (1996); Kim (1992, 2000); Lyre (2009).)

> "[O]n the traditional account of scientific reduction associated with Ernest Nagel, one theory reduces to a more basic theory when the former can be deduced from the latter by means of connecting principles that express property *identities*." (Endicott, 2005, 5 emph. added)

[13] I think that there are good grounds on which to resist thinking of the 'higher-level' theory as essentially richer than the 'lower-level' theory. For one thing, it presupposes the metaphysically inflated position that bridge-laws are metaphysically substantial (i.e. identities, nomic connections, or correlations) that I wish to avoid but I do not pursue this line of argument.

> "Ontological simplification is one main goal of theory reduction." (Esfeld and Sachse, 2007, 2)

As I have stressed, one of the positive features of Neo-Nagelian reduction is that it affords ontological simplification without recourse to the untempered metaphysics that riddles the literature on reduction and multiple realizability. Indeed, it seems to me that our imaginative ingenuity allows for ever more subtle metaphysical distinctions in this area[14] and, as such, the prospects of settling debates seem slim. This is not to say that it is not useful to invoke various kinds of distinctions - it often is - but taking all these distinctions metaphysically seriously is to the detriment of philosophical progress. Nonetheless, given how entrenched these debates are, it is necessary to examine the various arguments.

The plan for this section is as follows. In section 3.4.1, I recap how reduction is usually taken to afford ontological reduction - the *simplistic ontological simplification* thesis (SOS). The argument that multiple realizability prevents SOS is then considered. Next, I present a counter-argument to this, which is originally due to Lewis (1969) and taken up by Kim (1992)). Then in section 3.4.3 I will then consider the case of so-called 'radical' multiple realizability. Finally, in section 3.4.4 I will recap how to get ontological simplification without recourse to SOS, as previously shown in chapter 1.4.7 and show that this is not undermined in case of multiple realizability.

3.4.1 Simplistic Ontological Simplification

Simplistic Ontological Simplification (SOS) is the thesis that reduction affords ontological simplification via bridge-laws-cum-identities. That is, bridge-laws are taken to identify the properties of the higher-level theory with those of the lower-level theory. To understand how multiple realizability putatively undermines SOS it will be helpful initially to see how SOS operates in the non-MR case.

The starting point is the bridge-law itself. As before, I'll use the term 'realization' as a placeholder for the relation between the properties that predicates in bridge-laws denote.[15] In the non-MR case then, we shall speak of a single

[14]The list of metaphysical positions in this context is dazzling: property dualism, nomic correlations, brute correlations, strong supervenience, weak supervenience, token physicalism, non-reductive physicalism, mereological sums...

[15]This is intended to be entirely neutral and those who want to reserve the term 'realization' for a particular kind of relation distinct from, say, correlation, nomic connection, identity, supervenience etc, should introduce another term in its place.

lower-level property realizing a single higher-level property. It is important not to let the surface grammar of this last proposition mislead us: realization may turn out to be identity and that is not to be prejudiced by 'property' appearing twice in the previous sentence! What is not at issue here is where the relevant bridge-laws 'come from'. That is, I take as given, both in the MR and non-MR cases, that there is no disagreement about the *form* of the relevant bridge-laws but just about the *interpretation* of 'realization' in each case.

SOS in the non-MR case is motivated by parsimony. Given that the lower-level property realizes the higher-level one, parsimony dictates that they be identified. This seems to be an unassailable inference, *ceteris paribus*. Put another way, denying this would be to accept property dualism in the "best case scenario". It is easy to see why SOS is an attractive addendum to Nagelian reduction: one would get not only reductive explanation but also ontological simplification.

However, in case of MR-properties ontological simplification is not to be had, it is argued. Why? The argument is taken to be trivial: given that identity is a transitive relation, the higher-level property cannot be identical to its 'lower-level' realizers because the 'lower-level' properties are, *ex hypothesi*, not identical with each other. In terms of the earlier schemata, given a bridge-law like [4a] :

- $S_a \rightleftarrows P_a^1 \vee P_a^2 \vee \ldots P_a^n$

S_a cannot be identical with each of its lower-level realizers, P_a^is because, by transitivity of identity, the P_a^is would all then be identical to each other, in contradiction to the assumption that they are each distinct properties. Thus, MR-properties undermine ontological simplification. This argument is now discussed.

3.4.2 Local Simplistic Ontological Simplification

The earliest espousal of the previous argument against SOS can be found in Putnam (1967). Putnam's argument is framed in the context of psychology but nothing hangs on that other than, perhaps, the plausibility of the premise itself, i.e. that psychological properties like pain are multiply realized. In his review article, Lewis provides a defence of SOS as follows. Lewis takes Putnam to commit the brain-state theorist to the claim that:

> "all organisms in pain - be they men, mollusks, Martians, machines, or what have you - are in some single common nondisjunctive physical-chemical brain state. Given the diversity of organisms, that claim *is*

> incredible. But the brain-state theorist who makes it is a straw man. A reasonable brain-state theorist would anticipate that pain might well be one brain state in the case of men, and some other brain (or nonbrain) state in the case of mollusks. It might even be one brain state in the case of Putnam, another in the case of Lewis." (Lewis, 1969, 25 orig. emph.)

Lewis's point is this: there is no one kind of physical state, tokens of which are identical with tokens of pain; there are various kinds of physical states, tokens of which are identical with tokens of pain:

> "The seeming contradiction (one thing identical to two things) vanishes once we notice the tacit relativity to context in one term of the identities." (*ibid.*)

On this view, the 'higher-level' property is to be identified with 'lower-level' property specific to a particular domain. So, pain is identical with some neurological property in humans, and identical with some other neurological property in molluscs, and with something else entirely in say a robot, we can imagine. This view is accepted by some, cf. Kim (1992), Marras (2002), Sober (1999), Sklar (1993). The problem that the anti-SOSist finds with Lewis' position is that it forgoes the unity of the 'higher-level' property: In what sense can one still say that the 'higher-level' property is a single property?

This stance, viz. forgoing the unity of the 'higher-level' property, is precisely what Kim (1992, 1995) has advocated. It was discovered in the nineteenth century that the ornamental stone called 'jade' refers to two distinct metamorphic rocks, *nepthrite* and *jadeite*. This is a case of multiple realizability *par excellence*, Kim contends. Kim argues that 'jade' simply does not constitute a single property but is an abbreviation for the disjunction of 'nepthrite' and 'jadeite'.

> "Now, we put the following question to Fodor and like-minded philosophers: If pain is nomically equivalent to N, the property claimed to be wildly disjunctive and obviously nonnomic, why isn't pain itself equally heterogeneous and nonnomic as a kind? Why isn't pain's relationship to its realization bases, Nh, Nr, and Nm analogous to jade's relationship to jadeite and nephrite? If jade turns out to be nonnomic on account of its dual "realizations" in distinct microstructures, why

> doesn't the same fate befall pain? After all, the group of actual and nomologically possible realizations of pain, as they are described by the [multiple realizability] enthusiasts with such imagination, is far more motley than the two chemical kinds comprising jade." (Kim, 1992, 10)

A point of clarification is needed before considering this argument: what it is for *a property* to be *'nonnomic'* is not obvious. I think what Kim means here is that the property is nonnomic in the sense that it does not appear in any laws - it is its realizations that appear in laws, e.g. there aren't laws involving jade rather there are laws involving nephrite and jadeite. But, and this is Kim's point, if that's correct of jade, then it is also correct of pain - pain is thus an 'equally heterogeneous and nonnomic' (*ibid.*) kind. Kim's point strikes me as a salient one, but the emphasis on nomicity is detracting from it. To better see what is at stake here, suppose that, à la Lewis, one accepts domain-specific property identities - why then think that the higher-level property is a single property at all? Kim's point, as I understand him, is that the burden of proof is on the anti-SOSist's (though of course he does not use this moniker) shoulders to show why it is. One may grant that one can *name* the disjunction of properties just as one can name the disjunction of nephrite and jadeite 'jade' but a name carries no ontological significance.[16]

Whilst my intuitions are in line with Lewis' and Kim's, it is clear that an anti-SOSist will resist this argument. She can do this by arguing that the very fact that there are 'higher-level' laws ensures the singularity of the properties in terms of which they are couched. "Laws involving 'pain' are about *pain*!" As with all such disputes, there seems to be no simple way to settle it. Moreover, the anti-SOSist has a rejoinder based on 'radical' multiple realizability.

3.4.3 'Radical' MR

'Radical' multiple realizability putatively shows that are cases where domain-specific identities are purportedly not possible. This is best seen by example:

> "the intentional mental states we attribute to one another might turn out to be radically multiply realizable, at the neurobiological level of

[16] Clearly, if one considers the higher-level property to be nomically coextensive with the lower level property in each particular domain, a similar problem arises.

description, even in humans; indeed, even in *individual* human beings; indeed, *even in an individual human given the structure of his central nervous system at a single moment of his life*" (Horgan (1993, 308), as quoted in Bickle (1996, 123) Bickle's emphases)

In fact it is Bickle who presses this argument. He contends:

"[This] sort of multiple realizability is more troubling for reductionists, because a Lewis-inspired response won't work against it." (*ibid.*)

Why will domain-specific identities not work? Bickle continues:

"Relativising psychophysical reduction to specific domains now involves relativizing them to physical-chemical states of individuals at times... *Surely this much domain-specificity is inconsistent with the assumed generality of science.*" (Bickle, 1996, 124 emph. added)

Clearly more needs to be said about why this is inconsistent with the generality of science and, indeed, what "the generality of science" means.[17] But the intuition is clear: there may be cases where the a higher-level property is realized by a variety of different realizers in the same domain. This would then undercut domain-specific identities. (Again intuitions cut both ways.)

A putative example of such 'radical' multiple realizability is temperature. Lyre (2009) considers this, but comes to reject it. Here is how he puts it:

"Now, a certain macrostate of an ideal gas characterized by a particular value of the mean kinetic energy of its molecules may very well be instantiated by [a] gigantic number of microstates... all of which in general include different individual molecule velocities which nevertheless lead to the same mean value of the molecules' kinetic energy. Hence, temperature, a property of the macrostate of a gas, is multirealizable by a vast number of microstates. While this is in a sense a truly dramatic MR case, this does not at all mean that the reduction of temperature to individual molecule velocities or kinetic energies is blocked. The different microstates belonging to one macrostate may differ in various properties (e.g. the individual molecule velocities),

[17] These details are not forthcoming in Bickle's text, unfortunately.

> but as far as the relevant property is concerned Ű the mean kinetic energy of the molecules Ű these states are of one kind. Temperature laws may very well be reduced to laws about mean molecular kinetic energy despite the fact that the different microscopic realizations differ in various regards, since temperature laws do not quantify over the differing properties of the microscopic realizers, but only about the property in common." (Lyre, 2009, 8)

So what at first seemed like a case of 'radical' multiple realizability turns out *not* to be a case of multiple realizability at all! Yet, the anti-SOSist running the 'radical' line can just further re-entrench her position. Perhaps temperature is not a case of multiple realizability (let alone of the 'radical' kind) but there are surely others.

Shapiro (2000) proffers a general argument that purports to show that there are no such 'radical' MR-properties. To see it, first consider what Shapiro thinks counts as 'genuine' multiple realization:

> "Multiple realizations count truly as multiple realizations when they differ in causally relevant properties - in properties that make a difference to how they contribute to the capacity under investigation." Shapiro (2000, 644)

Shapiro's contention is that given this characterisation of multiple realizability, there are no 'radical' MR-properties i.e. that there are no properities which are realized by different properties in the same domain. Why this is so, is not entirely clear from Shapiro's paper. The implicit assumption seems to be that properties in the same domain are not different in their causally relevant properties. For example, reconsider Horgan's example above. An intentional mental state is a putative 'radical' MR-property because it is multiply realized in the same domain. i.e. there are various realizers of intentional mental states even in "one human being at a particular time" (*op. cit.*). But according to Shapiro (at least with the implicit assumption tacked on) intentional mental states will fail to be a MR-property because the putative realizers of it do not differ in their causally relevant properties.

There is much left wanting with Shapiro's argument. What is for realizers to differ in their causally relevant properties? This notion is incredibly vague. In

particular more needs to be said to ensure that the implicit assumption is not a gerrymandered non-sequitur!

So where do things stand with ontological simplification? As the dialectic above shows there is no consensus on the issue and prospects of settling the issue seem slim. As I stressed at the start of this section, it seems to me to be rather dubious to take any of this untempered metaphysics seriously. The Neo-Nagelian model of reduction does better. As I set out in chapter 1, taking bridge-laws to be metaphysically substantial (e.g. bridge-laws as identities, correlations, nomic connections, etc) is entirely misplaced. I substantiate this claim in the next chapter, chapter 4. Under the Neo-Nagelian model I am advocating, bridge-laws are not metaphysically substantial but are *coherence constraints*.

Whilst this allows me to cut through this metaphysical Gordian knot, where does it leave ontological simplification? Ontological simplification is clearly not to be had via SOS but then how does Neo-Nagelian reduction afford it? I turn to this question now.

3.4.4 Ontological Simplification Revisited

In this section I recap how Neo-Nagelian reduction can afford ontological simplification without SOS, that is, without taking bridge-laws to express identities, and I show that this method is unaffected by MR-properties. In broad strokes, my proposal is the following: provide an interpretation of Quine's meta-ontological position and show that, under that interpretation, the reduction of one theory to another results in a lack of ontological commitment to the ontology of the to-be-reduced theory. What follows is not a thoroughly detailed position but a programmatic suggestion. Nonetheless, I hope the reader finds it appealing.

Quine's meta-ontological position is well-known. It was set-out in Quine (1948). I shall adopt it here without further argument.[18] The slogan for it is equally well-known: "To be, is to be the value of a variable." (Quine, 1948). But this only tells half the story, for not any old variable will do. It is the variables

[18]Quine's meta-ontological position has, like any philosophical thesis, both its advocates and detractors. Famous, of course, is Quine's debate with Carnap on the question of meta-ontology. (Carnap (1950). Also, cf. Azzouni (1998).) There is no consensus about the correct meta-ontological position and settling the matter here is not my concern. Needless to say, I believe that there are good reasons for considering Quine's position although I will adopt it without argument. Those readers uneasy with this should conditionalise my claims about ontological simplification: *if* Quine's meta-ontology is essentially correct, *then* reduction affords ontological simplification.

in our best conceptual scheme which are the ones we need consider. Thus the slogan should be modified: "To be, is to be the value of a variable *in our best conceptual scheme*."

The pertinent question is now what our 'best conceptual scheme' is. Quine writes:

> "Our ontology is determined once we have fixed upon the over-all conceptual scheme which is to accommodate science in the broadest sense ['best conceptual scheme']; and the considerations which determine a reasonable construction of any part of that conceptual scheme, for example, the biological or the physical part, are not different in kind from the considerations which determine a reasonable construction of the whole." (Quine, 1948, 39)

As I said, the substantial work lies in explicating this, and doing so exhaustively falls beyond the scope of the present work. Hence what follows is programmatic. Quine himself does not set out in much detail what this amounts to but he does point us in the right direction:

> "[W]e adopt, at least insofar as we are reasonable, the simplest conceptual scheme into which the disordered fragments of raw experience can be fitted and arranged. " (*ibid.*)

I suggest the following development of Quine's suggestion[19]: the best conceptual scheme is one which strikes the best balance between simplicity and empirical adequacy but so that empirical adequacy is given priority. The 'priority' clause is to thought of like this: Sort conceptual schemes by empirical adequacy first, and then select of those which are equal in this respect, the simplest. This is, of course, not a precise procedure - a lot would need to be said about how to construct conceptual schemes and how to rank them in the relevant respects and so on - but I do think that the idea here is intuitively appealing and helpful. For even at this intuitive level, we can now see how to get ontological simplification out of reduction. If one theory, say T1, reduces to another, say T2, then any conceptual scheme in which T1 appears will not be as good as any conceptual scheme in which T2 appears. Why so? A successful Neo-Nagelian reduction requires that the reducing theory (T2) is empirically more adequate than the

[19]This is informed not only by the quoted passages but the rest of paper.

to-be-reduced theory (T1), because the laws of T1 are derived from those of T2 and counterfactual auxiliary assumptions. Notice that this is a conditional claim: it does not say that T2 is bound to be part of our best conceptual scheme but that if it is, then T1 won't be. The route to ontological simplification is now straightforward: if T1 is not in our best conceptual scheme then there is no ontological commitment to it. The properties that T1 speaks of may well be kept as useful fictions, but when the dust has settled we are not bound to take those properties as existent. In the next chapter, I shall show how this plays out for the temperature.

The point in recapping the Neo-Nagelian route to ontological simplfication is this: all of the above is unaffacted by multiple realizability. And this is an important result, I claim, for, as the forgoing discussion showed, it is specifically with repsect to ontological simplification that multiple realizability has most pinch.

3.5 Chapter Summary

In this chapter I have considered various arguments 'from' multiple realizability. First, it was shown that Fodor's orginal argument is ineffectual with respect the Neo-Nagelian model of reduction. I then considered whether multiple realizability poses a problem for the explanatory import that a NN reduction affords. I argued that it does not. Finally, I consider how multiple realizability impinges on the issue of ontological simplfication. I showed that, in so far as there are multiply realized properties in the 'higher-level' theory, it does pose a problem for ontological simplfication via SOS. However, this is not how NN reduction affords ontological simplfication and that latter is not affected by multiple realizability.

Chapter 4

Neo-Nagelian Reduction and CSM

4.1 Chapter 4 Introduction

In this chapter, I substantiate and exercise the Neo-Nagelian model of reduction. Recall the overall method advocated: to abstract a general model of reduction based on a rational reconstruction of the derivation of the Boyle-Charles law from the kinetic theory of gases. (cf. chapter 1) The Boyle-Charles law is just one of the *laws* of thermodynamics but what we are concerned with is, of course, inter-*theoretic* reduction. However, this seeming mismatch does not make for a substantive problem: the point is that if every law of a theory can be derived in the manner exemplified here, this will constitute the reduction of that theory to the reducing theory. I will sometimes speak of a law reducing to a theory but this is just a helpful abbreviation, in the above sense.

Whether every law in thermodynamics can be derived in the requisite way from statistical mechanics remains an open, but importantly, *theoretical* question. Were this to be achieved one would have a reduction of thermodynamics to statistical mechanics, and thereby a complete explanation of the empirical success of thermodynamics.[1]

To understand the Boyle-Charles law, and the questions pertaining to reduc-

[1] As discussed below, statistical mechanics is not a theory which has a canonical formulation; there is only a general rubric of attempts to account for thermodynamic phenomena via the molecular hypothesis and statistical assumptions. In this chapter we shall consider derivations from both the Gibbsian and Boltzmannian 'schools'.

tion here, a clear understanding of thermodynamics is needed. In section 4.2, I present an overview of thermodynamics and give a statement of the Boyle-Charles law. In section 4.3, I consider how it is derived. In section 4.3.1 I present the standard 'textbook' derivation of the Boyle-Charles law from the kinetic theory of gases. I then give a rational reconstruction of this derivation, indicating the features that constitute the Neo-Nagelian model of reduction, in section 4.3.2. Specifically, I shall focus on the notion of *warrant* for both the auxiliary assumptions and bridge-laws used. In the Neo-Nagelian account of reduction bridge-laws are a particular kind of theoretical stipulation, viz. *coherence constraints* and this will bare out in this section. This completes the presentation and defence of the Neo-Nagelian model.

In section 4.4, I go on to systematically show that bridge-laws *qua* identities is an untenable, indeed misconceived, position. Thus I repudiate the dominant position apropos bridge-laws.

The aim for the rest of the chapter is then to apply the Neo-Nagelian model of reduction. In section 4.6, I examine the derivation of the Boyle-Charles law from Gibbsian statistical mechanics. I show it to be unsuccessful. In section 4.7, I examine recent work on the derivation of Second Law of thermodynamics from Boltzmannian statistical mechanics. This, I show, is partially successful and promising.

4.2 Thermodynamics

Thermodynamics is one of the great successes of Nineteenth Century 'classical physics'.[2] In thermodynamics, systems are characterised in terms of various thermodynamic properties. Most textbooks on thermodynamics start with the properties of pressure, volume and temperature. A typical way of introducing them, in the first instance, is along *phenomenological* lines. That is, they are properties which are said to be 'observable' in the sense that they are detectable by human senses and for which there are measuring devices. For example, most discussions of temperature start with 'felt hotness and coldness' and proceed to operationalise it (cf. Zemansky and Dittman (1981, 10)). Thus, the sensation

[2]The start of thermodynamics as a discipline is usually attributed to Otto von Guericke, of vacuum pump fame, all the way back in 1650. Boyle and Hooke, both key figures, also worked from in the Seventeenth Century but it did not emerge as an individuated, independent theory until the 1820s, with the work of Carnot. The shift from the notion of heat as a substance to heat as a form of energy is also integral. For a detailed history see Müller (2007).

is used to create instruments that reproduce the *felt* ordinal ranking of various systems. These are what Chang calls 'thermoscopes' Chang (2007). These were then developed into thermometers through a calibration of a fixed scale. The measuring of various properties does not yet constitute a thermodynamic theory of course. What does, is the relation between the various properties that are ascribed to systems. It is the relationship between the properties posited as laws that constitute thermodynamics. Roughly speaking, there are two kinds of laws in thermodynamics: *Axiomatic* laws (namely the 'Zeroth', 'First', and 'Second') and *Constitutive* laws (cf. Zemansky and Dittman (1981) and Sklar (1993)). The former hold for every thermodynamic system. In fact, they define thermodynamics: any system for which they do not hold fails to be a thermodynamic system. The *Axiomatic* laws also serve to *define* the central properties of the theory. The constitutive laws are the phenomenological laws of the theory, and hold, to varying degrees of accuracy for a subset of thermodynamic systems. (e.g. there are laws that hold for certain kinds of gas, laws that hold for solids etc). This distinction will become clearer below.[3]

4.2.1 Axiomatic Laws

The cornerstone for thermodynamics is the notion of equilibrium. Intuitively, thermodynamic equilibrium is that state of the system for which the thermodynamic properties do not change over time.[4] Again intuitively it seems, then, that one can empirically check whether a given system is in equilibrium, using the various thermodynamic measuring instruments. For example, a gas could be kept at a fixed volume and its temperature and pressure could be checked using thermometers and pressure gauges to see whether they are unchanging. However, formally speaking the thermodynamic properties that these instruments are taken to measure are only defined for systems that are in equilibrium. The way out of this circularity is to take equilibrium to be a primitive of the theory.[5]

Equilibrium appears as a primitive in the Zeroth Law of Thermodynamics:

[3]Most of the technical work in the next three sections follows the discussion in Zemansky and Dittman (1981).

[4]More strictly: thermodynamic equilibrium is that state of a system when the mechanical, chemical and thermodynamic properties do not change over time, but we shall not be concerned with this detail. Throughout the rest of the discussion *equilibrium* is intended as thermal equilibrium.

[5]The circularity of introducing equilibrium in this way is often glossed over. Cf. Pippard (1957), for example.

> *Zeroth Law of Thermodynamics*: Consider two systems which are adiabatically separated. If the two systems are *each* in equilibrium with a third, then they are in equilibrium with each other.[6]

Temperature, T, is then to be introduced into the theory via the Zeroth law.

> "The systems themselves, in these states, may be said to possess a property that ensures their being in thermal equilibrium with one another. We call this property *temperature*. *The temperature of a system is a property that determines whether or not a system is in thermal equilibrium.*" (Zemansky and Dittman, 1981, 11 orig. emph.)[7]

Temperature is that property which systems in equilibrium with each other have in common. That is, one defines 'temperature' as such and it is a posit of thermodynamics that this term refers to a real property of thermodynamic systems.[8] It is also clear that temperature is an intensive property: just consider an arbitrary partition of the system into subsystems. These subsystems must be in equilibrium with each other, and hence have the same temperature.

It is important to the rest of the discussion to note that temperature is defined only in equilibrium. Strictly speaking, a system outside of equilibrium does not have a temperature.[9] The rationale one finds in textbooks for such an axiomatisation (i.e. for temperature being defined only in equilibrium) is an interesting

[6]This is the statement found in Zemansky and Dittman (1981, 10).

[7]Statements to the effect that the Zeroth law (allows one to) defines temperature are found in all textbooks on thermodynamics. Cf. e.g. Hecht (1998), or Pippard (1957).

[8]In fact, as it stands this textbook presentation is rather imprecise. However, no better ones have been found and for the present purposes the axiomatic foundations of thermodynamics do not matter all that much. A few comments about the shortcomings of the above presentation are in order, however. First, the Zeroth Law as stated above does not specify an equivalence relation; it only specifies a transitivity relation. Second, there is no distinction drawn between states being in equilibrium and one state being in equilibrium with another. Clearly these are not the same property: one is a property of an individual system and the other is a relational property between two (or more) systems. Third, the notion of adiabatic separation is not explicated. The first point is most important. Denote 'in equilibrium with' by $\cong$. To introduce temperature as this equivalence relation, one needs not only transitivity, as per the above, i.e. $(A_1 \cong A_2)$ & $(A_2 \cong A_3) \implies (A_1 \cong A_3)$, but one also needs that this relation be both reflexive and symmetric, i.e. $A_1 \cong A_1$ and $(A_1 \cong A_2) \equiv (A_2 \cong A_1)$. As regards the second point, it may be possible to define 'in equilibrium with' in terms of the equilibrium state - this has been suggested to me by Jos Uffink. As regards the third point, two systems are adiabatically separated when they cannot exchange heat.

[9]At least not in the thermodynamic sense. Central to the coming discussion is whether one can identify temperature with a statistical mechanical property and/or *extend* the concept of temperature 'beyond' equilibrium. cf. section 4.4.

one: outside of equilibrium the system as *a whole* does not have a *temperature*.[10] The point is that once temperature has been operationalized (via thermometers, say) were one to measure the temperature of the system one would not give a *uniform* result. To illustrate, consider a gas outside of equilibrium: were one to take thermometer readings in different regions of the gas, one would get different results. Thus one cannot assign the gas *as a whole* one value for temperature.[11]

Thermodynamic systems are not characterised solely by temperature. Certain mechanical properties are taken over from classical mechanics. Two are central: pressure and volume. The volume of a system is the amount of space that it occupies. The pressure that a system exerts on the walls of its container (or on the surface with which it is in contact, in the case of a solid), is equal to the force per area. For the coming discussion it will also be helpful to introduce the notion of the *equation of state*. The equation of state for a particular system is a specific functional relationship between the thermodynamic properties of that system in equilibrium.[12]

The 'First Law of Thermodynamics' encodes the idea that heat is a form of energy and expresses energy conservation.[13]

> *First Law of Thermodynamics.* The change in the internal energy, ΔU, of an isolated system is equal to the work done on the system, ΔW, and the heat gained, ΔQ.

The notion of mechanical work is taken over directly from Netwonian mechanics. Heat is *energy in transit* and is not straightforwardly directly measured. Heat, Q, is, however, related to temperature, T as follows.

$$\Delta T = \Delta Q / C \tag{4.1}$$

[10] Notice the significance of the italicization: 'temperature' refers to a theoretical concept, as defined above; '*temperature*' refers to a (supposed) property in the world.

[11] Whether this rationale is convincing is not something I discuss further. Here I am just presenting the standard modern way that thermodynamics is axiomatized.

[12] Saying that the system is 'in equilibrium' is strictly speaking superfluous for the ascription of thermodynamic properties presupposes equilibrium.

[13] As is well known, previously the so-called 'caloric theory' construed heat as a substance. Or slightly more carefully put, certain experiments which were accounted for by the positing of caloric were subsequently accounted for by heat instead. Moreover, certain experiments were seen to be incompatible with the 'heat as substance' view point. I do not discuss the relation between the caloric theory and modern thermodynamics. A compelling overview and collection of the seminal works, in particular the work of Celius and Joule, that led to the adoption of 'heat as energy' view. It was this move away from 'heat as substance' to 'heat as energy' which ushered in modern thermodynamics. cf. Brush (1986).

where ΔT is the change in the system's temperature, ΔQ is how much heat the system has gained/lost, and C is the heat capacity of the specific system, which depends on the process the system undergoes during the heat transfer.[14] This connection between the heat gained or lost by the system and the change in its temperature will be of significance in the rational reconstruction of the derivation of the Boyle-Charles law from the kinetic theory of gases.

The Second Law of thermodynamics is the most widely discussed and controversial of the axiomatic laws. There is no canonical formulation of the Second Law. The most comprehensive philosophical work on the issue is due to Uffink (2001), where Uffink identifies eleven distinct versions of the Second Law. (cf. Uffink (2001) 91) Really, there are various ways to formalise a general *tendency* pertaining to systems which naturally' fall under the remit of thermodynamics. As Pippard puts its:

> "[T]here is a certain tendency for changes to occur preferentially in one direction rather than for either direction to be equally probable. For example, we have taken it as a basic assumption, in accord with observation, that systems left to themselves tend towards a well-defined state of equilibrium. It is not observed that a reversion to the original non-equilibrium state occurs... the idea that there is a preferred direction for change has been perhaps most clearly expression [in terms of] *hotter* and *colder*. There is an unmistakable tendency for heat to flow from a body of higher temperature to one of lower temperature rather than for either direction of low to occur spontaneously. The second law of thermodynamics is little more than a generalization of these elementary observations. In essence it states that there is no process devisable whereby the natural tendency of a heat to flow from higher to lower temperatures may be systematically reversed." (Pippard, 1957, 29)

What is remarkable is that this tendency holds good for a vast variety of systems. It is this tendency which is to be encoded into thermodynamics via the Second Law. The axiomatic approach is to stipulate the following:

[14] One cannot speak of a *system's* heat capacity *simpliciter*, but rather the heat capacity of the system under a certain kind of process. Specifically: one might consider the heat capacity of a system under constant pressure, C_P: $C_P = \left(\frac{\partial dQ}{\partial dT}\right)_P$ or the heat capacity for fixed volume, C_V: $C_V = \left(\frac{\partial dQ}{\partial dT}\right)_V$. (cf. Zemansky and Dittman (1981, 84).)

Second Law of Thermodynamics. The entropy, S_{TD}, of an isolated system cannot decrease: $\Delta S_{TD} \geq 0$

As aforementioned, in modern expositions of thermodynamics the axiomatic laws are used to *define* the properties posited by the theory. Thus, entropy *is* that property which cannot decrease in isolated systems.[15] However, as stated this remains entirely abstract, and the law cannot be empirically motivated. It is motivated as follows.

Reconsider an arbitrary thermodynamic system. A transformation of such a system is a change in the state of the system. Consider two states of a *system*, E_1 and E_2. There is a quasistatic transformation of the system from E_1 to E_2 just in case the system remains in an *equilibrium State* during the transformation. That is, just in case all the intermediary states of the system are *equilibrium States*. A process is a series of transformations. A process is *cyclic* just in case any transformation of a *system* in state E_1, results in a final state of the system E_1. A process is reversible just in case it can be exactly reversed by an infinitesimal change in the external conditions. If a process is both cyclical and reversible it leaves the system and its environment unchanged. With this in place we can discuss the origins of the Second Law.

Originally Calsius formulated the Second Law of thermodynamics in terms of the work an engine can do:

> "It is impossible to devise an engine which, working in a cycle, shall produce no effect other than the transfer of heater from a colder to a hotter body." (Zemansky and Dittman, 1981, 187)

The Kelvin formulation is:

> "It is impossible to devise an engine which, working in a cycle, shall produce no effect other than the extraction of heat from a reservoir and the performance of an equal amount of mechanical work." (Zemansky and Dittman, 1981, 187)

How do these statements relate to the claim that the entropy of an isolated system cannot decrease? Consider a cycle consisting only of reversible processes. For some such cycle[16]:

[15] Notice this is the *thermodynamic* entropy. A system is isolated iff it is not in thermal contact with any other system and no work is being done on it.

[16] cf. (Dugdale, 1996) for details.

$$\oint \frac{dQ}{T} = O \tag{4.2}$$

This yields:

$$\int_{E_2}^{E_1} \frac{dQ}{T} = -\int_{E_1}^{E_2} \frac{dQ}{T} \tag{4.3}$$

for routes defined by the cyclic path, as per equation 4.2, from E_1 to E_2 and from E_2 to E_1. Hence the value of the integrals only depends on the beginning and the end point. This equation, equation 4.3, is used to *define* the difference in entropy between states of the system, E_1 and E_2:

$$S_{TD}(E_2) := S_{TD}(E_1) + \int_{E_1}^{E_2} \frac{dQ}{T} \tag{4.4}$$

where the change of state from E_1 to E_2 is reversible. Notice that the states we picked are arbitrary and as such this defines the entropy for every possible (equilibrium) state of the system.

However, reversible changes are an idealisation; changes for real systems are in general *irreversible*. What is of interest is change of the entropy under generic adiabatic irreversible changes, i.e. irreversible transformation in which there is no heat gained or lost. Given the statements of the Second Law, as per Kelvin or Calsius above, it follows that the entropy of a system cannot decrease.[17]

Suppose a system is in state E_1 and undergoes an arbitrary *irreversible* adiabatic transformation to E_2. To see that $S_{TD}(E_2)$ is greater than $S_{TD}(E_1)$, one shows that any cycle to take the system back to E_1 from E_2 results in the requisite inequality. By placing the system in contact with a heat reservoir at, say, temperature T_0, the system, in E_2, undergoes a reversible adiabatic change to E_3. Suppose now that the system undergoes an isothermal change from E_3 to E_4, such that the entropy of the system in E_4 is the same as what it started in, i.e. $S_{TD}(E_1) = S_{TD}(E_4)$. To affect the change from E_3 to E_4, either some heat, Q, must be added or taken away from the system. Finally the cycle is completed by taking the system from E_4 to E_1 by an adiabatic reversible change.

From the First Law it follows that the change in the change in the internal energy of this system is zero, i.e. $W + Q = 0$. There are two possible ways to satisfy this condition: either the system absorbs heat, $(+)Q$, and does work,

[17] Here I closely follow Dugdale (1996).

$(-)W$, or the system is worked upon, $(+)W$ and looses heat, $(-)Q$. The former possibility is ruled out however, for it would contravene the Second Law, as per both Clausius' and Kelvin's statements above:

> "If the first were possible we should have a cyclic process which did nothing else but produce work from a single temperature source at T_0. This work could then obviously be used in some suitable irreversible process to deliver an equivalent amount of heat to another body at any temperature, even one above T_0." (Dugdale, 1996, 61)

Thus, the system must *lose* heat, $(-)Q$, in going from E_3 to E_4. And this loss of heat corresponds to a decrease in entropy as per $\Delta S = Q/T_0$. Finally, as $S_{TD}(E_1) = S_{TD}(E_4)$, and $S_{TD}(E_2) = S_{TD}(E_3)$, it follows that:

$$S_{TD}(E_2) \geq S_{TD}(E_1) \tag{4.5}$$

It usually taken to be the case that the Second Law 'drives' the system towards equilibrium. However this is not quite right: the time-asymmetry of thermodynamics is more deeply ingrained into the theory. Brown and Uffink (2001) argue that equilibrium is itself a time-asymmetric notion:

> "[I]n thermodynamics the tendency of systems to approach equilibrium is logically prior to the Second Law ... The spontaneous motion towards equilibrium is time-asymmetric because of what equilibrium states are: once attained no spontaneous departure from them is possible without intervention from the environment. The equilibrium state in thermodynamics is itself a time-asymmetric notion." (Brown and Uffink, 2001, 527-528)

Following Brown and Uffink (2001) then one ought to introduce, what they call, the 'Minus First Law':

> *Minus First Law of Thermodynamics*: An isolated system in an arbitrary initial state within a finite fixed volume will spontaneously attain a unique state of equilibrium. (cf. Brown and Uffink (2001, 528-529))

It is really the Minus First Law which drives systems towards equilibrium. However, I shall speak of the Second Law encoding this aysmmetry as this both simpler and congruent with the majority of the literature.

This completes the presentation of the axiomatic laws of thermodynamics.[18] I now turn to the so-called constitutive laws.

4.2.2 Constitutive Laws

The constitutive laws of theory are the specific phenomenological laws. The constitutive laws are an application of the axiomatic laws to particular kinds of systems. Equations of state are laws of this kind. It is the constitutive laws that are the subject of experiment and they form the bulk of the empirical content of the theory. For example there are different laws for substances in different phases (e.g. gases, liquids and solids); there are laws for different (chemical) kinds of substances, and so forth. An important constitutive law is the Boyle-Charles law. It is the derivation of this law from the kinetic theory of gases that underpins the general model of reduction I advocating, viz. NN reduction.

The Boyle-Charles law states that the pressure, P and volume, V, of an ideal gas is directly proportional to its temperature, T:

$$PV = cT \tag{4.6}$$

where c is a constant.
It is worth re-emphasizing the difference in the role played by the constitutive laws and the axiomatic laws. In contrast to the axiomatic laws, the Boyle-Charles law does not define the thermodynamic properties. *Formally* one can solve the Boyle-Charles law for temperature resulting in the following equation:

$$T = \frac{PV}{c} \tag{4.7}$$

However, this is never done. Rather, the 'strictly' thermodynamic concepts, such as heat, entropy and temperature, are 'fixed by' the axiomatic laws of theory alone. (And the others are 'carried over' from mechanics.) Constitutive laws express the functional relationship between the properties that the concepts defined by the axiomatic laws (are taken to) denote for various specific kinds of systems.

In standard textbook presentations of this law it is stated that it holds for 'an ideal gas' (cf. Pippard (1957); Huang (1987); Hecht (1998); Zemansky and Dittman (1981)) What does this mean? It is important to distinguish an 'ideal

[18] One also sometimes speak of a Third Law of thermodynamics but it is not relevant to the rest of the discussion. In fact, it is controversial whether it is a law at all. cf. Frigg (2008). I do not pursue this point.

gas' in the thermodynamic sense from that in statistical mechanics. *In the thermodynamic sense, an ideal gas is defined as one for which this functional relationship holds exactly.*[19] As a matter of experimental fact, there are no real gases for which the law holds exactly. Rather, the law becomes more empirically adequate at lower pressures for real gases. (Which is just to say that as one lowers the pressure the error between what the law predicts and the outcome of the relevant measurements decreases. Conversely, the error increases at higher pressures.) Thus, a real gas could be said to *approximate* an ideal gas when it is sufficiently diffuse. (Moreover, at lower pressures the experimental error may be smaller than the 'theoretical' error.) Also, the law holds more or less accurately for different kinds of gases.

I now turn to the standard 'textbook' derivation of it from the kinetic theory of gases.

4.3 The Kinetic Theory of Gases

The so-called 'kinetic theory of gases' takes an ambiguous place in the taxonomy of physical theories. Contemporary textbook presentations of statistical mechanics tend to distinguish it from statistical mechanics 'proper'. (cf. Hecht (1998); Reiss (1996)). Although the fact that kinetic theory is covered at all in pretty much every textbook on statistical mechanics is telling! Why is it not considered part of statistical mechanics 'proper'? Textbook presentations tend to focus on the Gibbsian framework which is antithetical in its approach to the kinetic theory. (The Gibbsian framework derives its results by positing 'ensembles' and these are not used in the kinetic approach at all. More about this below.) However, statistical mechanics does not have a canonical formulation and there are 'schools' with differing methods and ideologies from the Gibbsian. (cf. Sklar (1993); Uffink (2001); Frigg (2008)) As such it would be contrived to insist that kinetic theory of gases is not part of statistical mechanics 'proper'. Historically, kinetic theory and statistical mechanics lay on a continuum, falling under the general rubric of attempts to account for thermodynamic phenomena via the molecular hypothesis and statistical assumptions. This is how I shall treat it too.

[19] In this sense the Boyle-Charles law *defines* an ideal gas in the thermodynamic sense.

4.3.1 Deriving the Boyle-Charles Law

What I present here is the standard derivation of the Boyle-Charles law as it appears in practically every textbook on statistical mechanics and/or thermodynamics. In the next section, section 4.3.2, we shall look at this with a more philosophical eye. Consider an isolated gas in a container of volume, V. Characterise this gas as consisting of particles obeying the Newtonian mechanics. One then makes the following assumptions. [20]

- The gas consists of a very large number, N, of particles.
- The particles are spheres of negligible volume (with respect to V) and of a fixed mass, M. (In this sense they are taken to be to like very small hard spheres.)
- For a small integrable volume dV, the number of particles contained within it, dN is still very large.
- The particles interact perfectly elastically with the each other and the walls of the container and there are no forces acting between them.

The motion of each particle is represented by a velocity vector, $\vec{v}$. The particles have various velocities in various directions. Now specify a velocity distribution, $f(\vec{v})$ such that there is no preferred direction for the velocity of the particles for all velocity vectors, $-\infty \geq \vec{v} \geq \infty$.

In Newtonian mechanics the pressure acting on a surface is defined as the ratio of force exerted perpendicularly upon it to its area. That is: $P = F_A/A$. Now consider one internal wall of the container in the x-y plane. What can be shown from the above assumptions is that the pressure acting upon this wall due to the particles colliding with it is:

$$P = \frac{M\ N}{V} \int_{-\infty}^{\infty} f(\vec{v}) v_z^2 d^3v \tag{4.8}$$

The integral in this equation is naturally defined as the *average* of the square of the velocity of the particles in the z-direction. Thus

$$P = \frac{M\ N}{V} \langle v_z^2 \rangle \tag{4.9}$$

[20] These are usually called 'simplifying assumptions' in physics textbooks.

Given the assumption that there is no special direction to the particle's velocities, it follows that $\langle v_z^2 \rangle = \langle v_x^2 \rangle = \langle v_y^2 \rangle = 1/3 \langle \vec{v}^2 \rangle$ from which it follows that:

$$P = \frac{MN}{3V} \langle \vec{v}^2 \rangle \tag{4.10}$$

Again from Newtionian mechanics, the kinetic energy of a particle, E_{kin} is equal half the square of its velocity times its mass: $E_{kin} = mv_i^2/2$. Thus, the average (or 'mean') kinetic energy of the entire gas is equal to $\langle E_{kin} \rangle = m \langle \vec{v}^2 \rangle / 2$. Thus one can now formulate an expression for the pressure exerted on the container by the gas in terms of the average kinetic energy of the particles:

$$P = \frac{2N}{3V} \langle E_{kin} \rangle \tag{4.11}$$

Rearranging this equation and reading it in the 'constitutive mode' (i.e. taking it specify the functional relations between the properties rather than defining any of them), we have an expression that asserts the direct proportionality of the pressure and volume of a gas to the mean kinetic energy of the particles that constitute it:

$$PV = \frac{2N}{3} \langle E_{kin} \rangle \tag{4.12}$$

In textbook presentations one proceeds now by 'directly comparing' this equation with the Boyle-Charles law. i.e. one can substitute the right-hand-side of equation 4.6 into the left-hand-side of equation 4.12 and solving for temperature yields:

$$T = \frac{2N}{3c} \langle E_{kin} \rangle \tag{4.13}$$

Recall, that the stated goal of kinetic theory of gases is to derive the Boyle-Charles law. Thus equations 4.12 and 4.13 are used to derive equation 4.6. This completes the standard derivation of the Boyle-Charles law from kinetic theory.

4.3.2 A Rational Reconstruction

One of the main tasks of statistical mechanics is to derive the laws of thermodynamics and, *ipso facto*, explain their empirical success. However, justice cannot be done to the whole of thermodynamics and statistical mechanics in this sense: the laws of thermodynamics, once one takes account of all the constitutive laws,

are simply too numerous. Instead, I have proposed to take the derivation of the Boyle-Charles law as the exemplifying case.

It is important to re-emphasize that this - the derivation of the law - is something that is part and parcel of statistical mechanics and not just philosophers' fancy! A derivation of the Boyle-Charles can be found in every treatise and textbook on statistical mechanics.[21]

So now let us look at the derivation and examine the way in which it is explanatory. The reducing theory here is the kinetic theory of gases. The 'metaphysical picture' is one of the system consisting of N identical classical particles, of mass m and volume V. The particles have kinetic energy of various forms (translational, rotational, etc) there is also potential energy between them. At any one time each of the particles has a definite position and velocity, (they are traveling on definite trajectories, or colliding with another particle or the internal walls of the container). The entire system evolves deterministically according to the governing Hamiltonian. If we had a perfect (informationally complete) model of this system and were able to solve it, we would (amongst other things) be able to predict the entire future evolution of the system.

Call the derivation of the Boyle-Charles law in the above form the *reductive construction*. What does it consist in? A conjunction of the reducing theory and various auxiliary assumptions. Setting aside the purely formal ones, these auxiliary assumptions are counterfactual from the point of view of the reducing theory.

First consider the auxiliary assumption of the Limit kind (AA2): we assume that the number of particles is sufficiently large so that the number of particles contained within it a integrable volume, dV, is still very large. From a purely mathematical point of view, the number of particles would have to be infinite to afford the requisite integration.

The velocity distribution assumption is an Initial condition kind of auxiliary assumptions (AA4). It actually asserts two things: that the distribution of the velocities of the particles is independent of the magnitude of the velocities themselves (for example, ruling out the possibility that the fastest moving particles are moving in one direction and the slowest in another direction) *and* that there is no 'preferred' direction of movement for the particles (for example, ruling out the possibility all the particles are moving in the same direction). It is this as-

[21] Historically, too, this was of major significance: deriving this simple constitutive law was a kind of a litmus test for the viability of the 'molecular approach', cf., for example, Brush (1986)

sumption that allows one to take the pressure exerted on the X-Y plane wall to be proportional to exactly one third of the average velocity of the particles.

The final auxiliary assumptions here are the Idealisations and Dynamical assumptions (AA1 and AA3). It is assumed that the particles are point-like, and that they only have translational kinetic energy, and only interact perfectly elastically with one another and the walls of the container. (That the particles only have translational kinetic energy entails that they can only interact perfectly elastically, on the further assumption that the internal energy of the system is fixed.) Clearly it is crucial to the derivation that the sole form of energy that the particles can take is translational kinetic energy. It is these assumptions that yield equation 4.11. Bracket for the sake of argument any worries you may have about equation 4.13, just for a moment: Given this equation, viz. equation 4.13 we can derive the Boyle-Charles as above.

Have we explained the empirical success of the Boyle-Charles law? The derivation fits the DN model of course, but the pertinent question is whether there is *warrant*, in the sense introduced in chapter 1.4, for the auxiliary assumptions and the bridge-law. Without there being *warrant* for the them - that is, without the auxiliary assumptions being explanation supporting - the derivation is not explanatory. Now, in the ideal case one would have a derivation of the exact laws of the to-be-reduced theory from the reducing theory and auxiliary assumptions and bridge-laws which are maximally *warranted*. But the crucial thing to remember is that the *warrant* for the auxiliary assumptions and bridge-laws is not determined by whether the laws of the to-be-reduced theory are exactly derived.[22] These are two independent factors which *together* determine how good an explanation one has: auxiliary assumptions which fail to derive the laws of the to-be-reduced theory or least laws close to them, cannot be said to an explanation of them, irrespectively of how *warranted* they are, but conversely, deriving the laws, even if exactly, from a set of auxiliary assumptions for which there is no *warrant* simply does not count as an explanation either. So what counts as *warrant* here?

The *warrant* for AA2 is straightforward. The more mathematically rigorous a derivation is, the better an explanation based on it. (cf. Batterman (2000).) The right way to think about the *warrant* for AA4 is how 'special' it is: the more general, or the less specific, the assumption the more *warranted* it would be.

[22]The various models of reduction we have encountered, viz. Kemeny and Oppenheim, Nagelian and New Wave Reductionist, all suppose that whether a reduction is successful is determined by whether the laws of the to-be-reduced theory *are* exactly derived.

After all, the Boyle-Charles law holds (at least approximately, though to varying degrees) of a great variety of systems irrespectively of their initial conditions, or so it seems.

What about AA1 and AA3? The right way to think about *warrant* for these is not how counterfactual they are with respect to PW_R. In order to derive the Boyle-Charles law, we *need* counterfactual auxiliary assumptions. The auxiliary assumptions encode that, *from the point of view of the reducing theory* (and indeed from the point of view of the actual world), i.e. from the point of view of the 'metaphysical picture' of the kinetic theory of gases, the Boyle-Charles law is strictly speaking false. The behaviour of gases at the possible world at which the reducing theory is literally true, call it PW_R, is not in strict accordance of the Boyle-Charles law. This is true *ex hypothesi*, the possible world in which the Boyle-Charles laws is literally true, PW_{BC} is the possible world at which the kinetic theory of gases is true *and the auxiliary assumptions are true*, which is distinct from PW_R. It is then straightforward to see that the counterfactualness of these auxiliary assumptions cannot be a measure of the *warrant* for the auxiliary assumptions. The counterfactualness of them is a product of the (lack of) verisimilitude of the Boyle-Charles law with respect to PW_R. Rather, they are *warranted* in so far as making them less counterfactual yields empirically more adequate laws. For example, the assumption that the particles in the gas only have translational kinetic energy is *warranted* in so far as assuming, say, that they also have potential energy, yields empirically more accurate laws. And indeed, this is the case. Consider, for example, the Van der Waals equation. (cf. Hecht (1998)) Never mind the derivation; the crucial point for our purposes is that in assuming, as is done, that there is pair-wise interaction between the particles of the gas yields an empirically more accurate law than the Boyle-Charles law. That is, making the AA less counterfactual with respect to PW_R yields an empirically more accurate law at the actual world.[23] Given that the laws of the to-be-reduced theory are strictly speaking, false, it is not how counterfactual the auxiliary assumptions are, *per se*, which matters but rather whether they are counterfactual *in the right way*.[24] For instance, that in making the auxil-

[23]We are of course concerned with empirical adequacy and explanation at the actual world. Notice, by definition, that making the auxiliary assumptions less counterfactual with respect to PW_R yields more empirically adequate laws at PW_R.

[24]If the laws are true (i.e. empirically perfectly adequate) then, of course, the auxiliary assumptions cannot be counterfactual. This will be discussed in more detail when we consider the derivation of the Second Law in section 4.7.

iary assumption about particles only having kinetic energy less counterfactual by modeling them as having potential energy too, and in so doing deriving more empirically adequate laws, we are assuring ourselves that the original assumption was not just ad hoc. It makes sense from the point of view of the reducing theory; it is not just surreptitiously used to yield the right result.

Now let's consider the bridge-law. It will, no doubt, have been immediately observed that the above derivation involves a seeming circularity: the method of 'direct comparison' basically allows one to derive the temperature equation, equation 4.13, from equation 4.12 and the Boyle-Charles law. However, then the temperature equation, equation 4.13, in conjunction with equation 4.12 are used to derive the Boyle-Charles law. But isn't that just presupposing the very thing one wanted to derive? Not quite. Equation 4.13 is a theoretical stipulation. The question is whether this stipulation is explanation supporting, or whether it undermines the explanatory force of the derivation. As I show below, it is a particular kind of theoretical stipulation which satisfies *formal consistency* and *conceptual fit*. This is incorporated into the Neo-Nagelian account. I call such a stipulation is a bridge-law *qua coherence constraint*.

Recall how we proceeded: the 'metaphysical picture' of classical mechanics is supposed. This is the 'background supposition". In possible world semantics, we suppose that we are at the possible world in which the reducing theory, in this case classical mechanics, is true, PW_R. The gas is then modeled in the 'kinetic theoretic way', which is just to say that it is modeled as consisting of a very large number of hard spheres of negligible volume interacting perfectly elastically under the assumption of molecular chaos etc. These are the counterfactual auxiliary assumptions. (Notice they are counterfactual at PW_R.) We can then ask, as it were, what temperature would have to be, if the kinetic theoretic model is to be consistent with the Boyle-Charles law. This fixes the particular form of the bridge-law. Notice this is not to suppose that the Boyle-Charles law is true (or correct) - rather it is constraining what temperature would have to be if the Boyle-Charles law *were* true given that the kinetic theoretic model *is* true (or supposed true). To make the construction (that is, the kinetic theoretic model) consistent with the Boyle-Charles law, temperature needs to be related to mean kinetic energy as per equation 4.13.

Immediately we face the spurious reduction problem. (cf. chapters 1 and 2.) To illustrate what the problem is consider an imaginary case: one wants to derive

a macroeconomic law from quantum mechanics.[25] For the sake of argument say that there is some macroeconomic law which involves properties X1, X2, and X3, and posits that X1 and X2 are directly proportional to X3. To proceed in a manner similar to the above, imagine one supposes quantum mechanics and various auxiliary assumption, and then derives some equation of a functional form identical to the macroeconomic law. One could then proceed as with the Boyle-Charles law and 'directly compare' the two equations, viz. the equation expressing the macroeconomic law and the equation just derived, and find an expression relating X1, X2 and X3 and some quantum mechanical properties. One could then introduce these expressions as theoretical stipulations. Finally one could then derive the macroeconomic law from the conjunction of the 'reducing-theory', auxiliary assumptions and 'bridge-law'. Fancifulness aside, doesn't this show that formal consistency is just *too weak*? Quite. To avoid the spurious reduction problem a bridge-law needs to encode more than just consistency. That is, to ensure that this move is not too 'cheap' we must go beyond mere *formal consistency* to *coherence*, the difference lying with *conceptual fit*. In the macroeconomics to quantum mechanics case, there is no *conceptual fit*.

The important point to underline here is that a bridge-law *qua coherence constraint* requires more than just *formal consistency*: it has to make *conceptual sense*! In the macroeconomics case, one has *consistency* (the example was constructed as such) but there is no *conceptual fit* because there is no relation between the quantum mechanical properties and the macroeconomic ones. Hence this is a spurious reduction.

The bridge-law in the Boyle-Charles law is arrived at in a similar manner as the (imagined) macroeconomic-quantum one, i.e. by the requirement of consistency. Supposing that the kinetic theoretic model is true constrains with what temperature is equated on pain of consistency. The equation certainly does not 'drop out of the sky'. Once a gas is taken to consist of a swarm of particles as per the kinetic-theoretic construction (the *reductive construction*) then all temperature could be equated with is some sort of bulk property of the kinetic energy of the particles. *Consistency* entails the particular form of the bridge-law but this is not sufficient for reduction - we also need *conceptual fit*.

The bridge-law states that temperature is directly proportional to mean kinetic energy. In what sense is there *conceptual fit* for the bridge-law? Recall that

[25]Remarkably, this is not all that far fetched: "quantum finance' is an active research area! cf., for example, Chen (2004).

we are considering an ideal gas in equilibrium, with fixed temperature in a finite volume on which no work is done. For such a gas, any change in the system's temperature is (would be) directly proportional to heat gain (or loss), as per equation 4.1. Heat is a form of energy as is clear from the First Law: the change in a system's internal energy is equal to the work done on/by the system plus its the heat gain/loss. And for a system on which no work is done, the change to its internal energy can only be in terms of heat loss or gain. Of course, that temperature is *directly proportional to* one form of energy for a thermodynamic system, does not meant that temperature *is* a form of energy; temperature and heat are two distinct properties. Yet, that they are proportional shows the *conceptual fit* because the bridge-law too states that temperature is directly proportional to a form of energy, namely mean kinetic energy. It is also important to reiterate that in taking temperature to be directly proportional to mean kinetic energy, the bridge law is not, so to speak, leaving other forms of energy unaccounted for: the particles of an ideal gas *only have kinetic energy.*

It might be objected that this is a rather thin conception of *conceptual fit*: is showing that in thermodynamics temperature is directly proportional to a form of energy enough? The right response to this question is another: enough for what purpose? The purpose of requiring bridge-laws to have *conceptual fit* is to avoid the problem of spurious reductions. Showing that temperature is directly proportional to a form of energy is sufficient for this purpose. As argued in Chapter 1, there are no general necessary and sufficient conditions for *conceptual fit*, rather it is a context specific and textured notion. For any putative bridge-law, an argument needs to be made that the bridge-law is not just formally consistent, but, to avoid spurious reduction, that the properties being related in the bridge-law fit conceptually.

Bridge-laws have been, and continue to be, a source of much discussion in the philosophical literature. In chapter 1.3.2 I identified two related problems for Nagelian reduction *vis-à-vis* bridge-laws: Where-From and Status. Respectively, the problems are that on the Nagelian model these laws seem to 'drop out of thin air' and that the status of bridge-laws is mysterious at best. As there noted, the first problem is hardly recognised, let alone dealt with. As regards the latter, there is much disagreement. We saw that the possibilities considered in the literature are that they express conventions, semantic claims, or some sort of metaphysical relation. The latter is further specified: some argue that bridge-laws express brute

correlations between properties, others that they express nomic connections, and others still that they identify properties.

It is important to understand all this in a way that avoids it being misleading. First, it is important not to gloss over the auxiliary assumptions that are integral to the derivation. The bridge-law has that particular form because it is assumed that the particles can only have translational kinetic energy, have negligible volumes and so forth. In short, the bridge-law has that particular form because of the *very particulars* of the kinetic-theoretic model postulated. Note for example that it takes a different form if one assumes that the gas is diatomic rather monotonic. In this sense, the bridge-law indicates the particulars of the construction, in a way that would be missed were one to simply say that the temperature must be related to mean kinetic energy given the concept of temperature, heat and internal energy.[26] Now consider the question of the status of bridge-laws. It is clearly a theoretical stipulation. But it is a theoretical stipulation for which there is *warrant*! The *warrant* for a bridge-law comes from showing it be a *coherence constraint* i.e. showing it to satisfy both *formal consistency* and *conceptual fit*.

In the next section, I shall argue that treating bridge-laws as metaphysically substantive - as is the dominant position in the philosophical literature - is misguided.

4.4 Is Temperature Mean Kinetic Energy?

In this section, I consider and reject the consensus view about bridge-laws, namely that bridge-laws express some sort of metaphysically substantial relation between properties.[27] By a careful examination of the temperature-mean kinetic energy bridge-law, as per equation 4.13, I shall argue that it is misguided to think of bridge-laws in any metaphysically substantial way. This in turn motivates the interpretation of bridge-laws qua *coherence constraints* I am advocating.

What is meant by 'metaphysically substantial'? The metaphysically substantial interpretation of bridge-laws takes the predicates in bridge-laws to refer to *real* properties in the world and the bridge-law to express just what the relation between them is. The first option is *identity*: bridge-laws express the identity of

[26] When we come to discuss derivations of the 2nd Law of Thermodynamics, it will be seen that the notion of conceptual fit is not straightforward: how to show conceptual fit for thermodynamic and statistical mechanical entropies is more substantive and contentious matter.

[27] But that is not to say that there is consensus about which *specific* kind of relation bridge-laws express.

the referents of the two predicates. Second, *nomic correlation*: bridge-laws expresses that the referents of the two predicates are nomically correlated. (That is, they are correlated and this correlation is a law-like.) Third, *correlation*: bridge-laws express that the referents of the two predicates are correlated. (That is, they are correlated but this correlation is contingent not law-like.)[28] These three options do not exhaust the set of possible relations a bridge-law could express, but these are the possibilities which are considered in the literature.

In what follows I shall argue that interpreting bridge-laws as metaphysically substantial in this sense, is a misguided enterprise. In particular, I shall focus on the arguments for the *simplistic ontological simplification* (SOS) thesis - the thesis that bridge-laws express property identities. In section 4.4.1, I first examine SOS. I identify several arguments for SOS but find only one to be plausible. In section, 4.4.2 I show why this argument is not persuasive. In sections 4.4.2.1 through to 4.4.2.4, I consider other ways in which one might 'save' SOS and find these to be wanting too. However, in section 4.4.2.5 I go on to argue that one ought not to then conclude that temperature is a kind of emergent property in light of the failure of SOS. In section 4.4.3, I shall recapitulate the right way to think about how NN reduction can afford ontological simplification.

4.4.1 *A Priori* Arguments for Identity

If a bridge-law expresses the identity of the referents of two predicates then there is ontological simplification. On this much there is consensus. *What were thought to be two distinct properties are the same property.* Some of those who deny identities do so on empirical grounds. The argument is that *empirically* one can only establish the *correlation* of the properties and that therefore there is not sufficient reason for the identity claim but only the 'weaker' correlation claim. This is the view of put forward by Brandt and Kim (1967); Kim (1995).[29]

There is a precedence in the literature for conceding this line of reasoning even by those that advocate identities. (cf. for example, Hooker (1981); Marras (2002); Needham (2010) for critical discussion.) That is, it is conceded that identities are *empirically* indiscernible from correlations. In favour of identities, what is

[28] Supervenience, on my understanding, is a modal notion: one set of properties supervenes on another just in case there *can* be no change in the former without a change in the latter. In the context of bridge-laws, one can think of supervenience between sets of properties as a kind of correlation.

[29] Although in more recent work Kim has come to defend identities: Kim (2000).

usually argued is that there are extra-empirical reasons for them. Representative of this view are Ager et al:

> "Experimental evidence is not what decides between the stronger 'Temperature is identical with mean kinetic energy' and the weaker 'Temperature is correlated with mean kinetic energy.' We deny, however, that strictly experimental considerations are the only relevant considerations in the temperature and mean kinetic energy case. We have tried to bring out conceptual considerations by which we can reach identity instead of stopping with correlation ... [W]e say that it is partly the nature of the case and partly the intent of the strategy of reduction that makes us say that temperature *is identical* with mean kinetic energy. It is the experimental evidence which makes us say that 'Temperature is identical with mean kinetic energy' *is true*." (Ager et al., 1974, 128-129)

Ager et al assert that the identity claim is 'stronger' than the correlation claim. Just what that means is entirely obscure. In the logical sense, to say that A is 'stronger' than B is to say that A *implies* B. But it is not the case if two properties are identical that this *implies* that they are correlated. A conceptual prerequisite for there to be a correlation between two properties is that there be two of them! Identity and correlation are distinct metaphysical relations.[30]

Setting aside the idea of the relative strengths of the relations, what are Ager et al's arguments that temperature is identical to mean kinetic energy? Whilst it is not clear what exactly is meant by the 'nature of the case' and 'the intent of the strategy' (*ibid.*) one can glean two kinds of arguments for the identity claim by a close reading of the text. The kinds of arguments that Ager et al. put forward are also echoed by others. (*op. cit.*)

The first is a kind of local inference to the best explanation. Ager et al write:

> "The result of this identification is that we have gained new information about temperature, and it is this new information which enables us to explain or understand why the pressure, volume, and temperature of a gas are related as they are described in the empirical gas

[30]This particular stance as regards 'strength' is by no means isolated: that identity is a 'stronger' relation than ('mere') correlation is echoed throughout the literature on reduction. Nomic correlation usually takes an 'intermediate' position (i.e. 'stronger' than correlation and 'weaker' than identity). cf. Hooker (1981).

> equation. For example, since the temperature of a gas turns out to be the mean kinetic energy of gas molecules, we can explain why, at constant volume, increasing the temperature increases the pressure" (Ager et al., 1974, 124)

Whilst this is not explicitly an inference to the best explanation (IBE), it is clearly implied (especially when conjoined with the passages preceding the one given): the best explanation of the functional relation between temperature and pressure is that temperature is mean kinetic energy. There is, of course, much to be said about the status of IBEs and whether they are a tenable inference pattern. For the present purpose, I shall bracket these broader issues. There is a specific problem with use of an IBE in this context: why is identity considered the *best* explanation? It is not at all obvious that it is. A 'brute' correlation entails the same functional relationship and it is therefore *as explanatory apropos the explanadum*, namely the 'empirical gas equation' (the Boyle-Charles law). That is: in so far as what is to be explained is concerned, correlations - and this is crucial point - *if true* explain the functional relation of the thermodynamic properties as per the 'empirical gas law' just as well as identities. To make this vivid replace identities with the perfect correlation in the previous quote: "since the temperature of a gas turns out to be perfectly correlated with the mean kinetic energy of gas molecules, we can explain why, at constant volume, increasing the temperature increases the pressure". This is no less explanatory than the case where temperature is identical with mean kinetic energy with respect to why, at constant volume, increasing the temperature increases the pressure. So, whatever else one may think about IBEs in general, in this specific context they cannot be used to advocate bridge-laws *qua* identities over bridge-laws *qua* correlations.

A second kind of argument (and one which may be seen as a retort to the objection I lodged against the first) is that without identity "temperature would be an ontological dangler with respect to the kinetic theory." Ager et al. (1974, 123). I say a second *kind* of argument because there are two ways to take this point. One might read it as a prudential argument for avoiding an explanatory gap: *temperature better be identical to kinetic energy for otherwise an explanation of why the two are correlated would be required.* Call this the 'prudential argument'. Alternatively, one can read it as an argument from parsimony: *identifying the properties would be more parsimonious then correlating them, so by a principle of parsimony, they are identical.* Call this the 'parsimony argument'.

The prudential argument is widespread in the literature on reduction (cf. Sklar (1993); Kim (1995, 2000)).[31] Whilst the prudential argument establishes the utility of identities over correlations, it cannot establish the truth of identities. It does not follow from that the fact (if it is a fact) that it is in some sense *better* to have identities that it is true that bridge-laws express identities.

What of the parsimony argument? This too is widespread in the literature on reduction. If one accepts the principle of parsimony - Ockams's razor - the argument does speak in favour of identities over correlations. This is not an uncontroversial 'if' of course: there is much discussion about the tenability of Ockam's razor as a 'true' metaphysical principle.[32] For the present discussion, I will regard the parsimony argument as a *defeasible argument* for identities. That is, I propose to concede that the parsimony argument establishes identity bridge-laws - in the case in hand the identity of temperature and mean kinetic energy - over correlations (nomic or otherwise), *modulo defeating reasons*. However, in the next section I provide defeating reasons for identity.

4.4.2 Conceptual Arguments Against Identity

The first (and most obvious) argument against identifying temperature and mean kinetic energy is that for many systems it is false that the temperature is equal to the mean kinetic energy of the system as per equation 4.13. Equation 4.13 is simply false for solids, for example.

If the equation is false in some cases, then temperature cannot be identical with mean kinetic energy, at least not as expressed by this equation. This is because of the transitivity of identity.[33] Of course, once one has seen where the bridge-law 'comes from' this is to be expected: the bridge-law is a *coherence constraint* given the kinetic-theoretic construction of *an ideal gas*. That is: equation 4.13 has that particular form because of the auxiliary assumptions that went into constructing an ideal gas. But there have been several suggestions for how to 'save' bridge-laws-cum identities. I consider these in the following sections.

[31] What I have here called the prudential argument ought not to be confused with the prudential argument of Fodor's as characterised by Sober in the context of multiple realisability, as per chapter 3.3.

[32] Establishing *a priori* metaphysical principles via analysis facilitated by armchairs is something that I am hesitant to endorse although a thorough-going discussion of this falls beyond the scope of the present work.

[33] Note that *this* does not preclude identifying temperature with some form of mean energy but it cannot be identical to just kinetic energy, as per equation 4.13.

4.4.2.1 Local and Counterfactual Identity

How might one 'save' the identification? One might posit a restricted or 'local' identity i.e. 'temperature is identical with mean kinetic energy for ideal gases'. This will sound familiar of course: it is just what is advocated initially by Lewis (1969), and further elaborated and defended by Bickle et al, in response to the multiple realizability argument. (cf. chapter [3, ref]). In that context, the argument was that a multiply-realised property cannot be identified with its lower-level realizers for, *ex hypothesi*, the lower-level properties are not identical with each other. The response was to 'locally' identify the properties or, as Lewis puts it, restrict the identification to a particular context. Thus, schematically, one would then identify 'temperature for gases' with one lower-level property, 'temperature for liquids' with another lower-level property and so forth.

However, even a restricted identification of this sort is not possible. Why? It is not possible to identify temperature and mean kinetic energy for *ideal gases* because 1) if there are no ideal gases then one cannot identify its putative properties with others (for there is no property to identify!) and 2) *there are no ideal gases*.

I take it that the truth of the first premise is self-evident. As regards the second, we have already encountered it: An ideal gas, in the thermodynamic sense, *would be* a gas which obeys the Boyle-Charles law exactly but no real gases do; real gases approximate the ideal gas. More carefully put, the empirical values of the thermodynamic properties of (a restricted set of) real gases approximate the values given the Boyle-Charles law, and do so with increasing accuracy at lower pressures. But the bridge-law, equation 4.13, only holds for ideal gases. One could say something counterfactual: if *there were* ideal gases, then temperature *would be* identical with mean kinetic energy for them. Even if one can make sense of counterfactual identities of this sort, clearly temperature is not identical with mean kinetic energy for real gases.

At this point it might start to look like as if the bridge-law in question should be rejected as false! But to think so is to have been mislead. The bridge-law expresses the right functional relationship between temperature and mean kinetic energy for ideal gases. What one has to not lose sight of, is that the kinetic-theoretic construction (*reductive construction*) reflects that the Boyle-Charles law is strictly speaking false (and that there are no ideal gases, in the thermodynamic sense): gases *aren't* made up of tiny hard spheres that interact perfectly elastically

etc. The bridge-law is a *coherence constraint* on the kinetic-theoretic construction - *it is what needs to be assumed to make the kinetic theoretic construction coherent with the Boyle-Charles law*, recall. Of course, formal consistency is not enough, as I have stressed. There is conceptual fit: in thermodynamics temperature (via heat capacity) is the only form of energy that an isolated ideal gas, in thermodynamic sense, can take and the kinetic-theoretic construction is such that an ideal gas only has translational kinetic energy. The bridge-law expresses this is in a mathematic-physical form. All of this makes sense once one is clear what the reduction is *for*. It worth restating the crux of the matter here again. One is trying to account for (explain) the (not complete but partial) empirical success of the Boyle-Charles which is, strictly speaking, false. One supposes another theory - here kinetic theory of gases - which forms one of the explanans. Various other explanans are needed - these are the auxiliary assumptions and the bridge-law. The bridge-law is a particular kind of theoretical stipulation, namely a *coherence constraint*, which serves as an explanan in the derivation of the Boyle-Charles law.

4.4.2.2 Identification with 'Complex Mean Energy'

Pursuit of identification might run in another direction. Clearly what is problematic in the above is the fact that there are no ideal gases. But surely real gases do have a temperature and this should be identified with some statistical mechanical property. It might be suggested that for real gases, temperature be identical not with mean translational kinetic energy but with some more complex lower-level property. Identify temperature with, say, mean translational, vibrational, rotational and potential energy of the particles, it might be implored. This suggestion is a more prosaic restatement of the point made above that temperature for a gas is directly proportional to its internal energy which must somehow be related to the different forms of energy that the particles can take, once a gas has been identified as a collection of particles. But for this not to be mere hand-waving, it needs to find rigorous mathematico-physical expression: what *exactly* are we to identify temperature with? The problem is that some such mathematico-physical expression is simply not forthcoming.

For the sake of argument, let us suppose a more sophisticated model is constructed which yielded (in the same way as the above) an expression equating temperature to a complex mean of the different forms of energy that the particles are modeled to possess. Suppose further that law being derived, for which this

expression acts as a bridge-law, holds exactly for real gases for a wide range of temperatures. But now we could no longer use this new bridge-law to derive the Boyle-Charles law! It would give us the wrong result; we would be deriving a 'corrected' Boyle-Charles law. That is all well and good in and of itself but the aim was to explain the empirical success of the Boyle-Charles law and not some corrected version of it.

Even setting this aside, however, would we be justified in identifying the properties via this (hypothetical) equation? In this case, it *seems* that an identity is possible. Certainly the previous argument - that it would be blocked because there are no ideal gases - does not have purchase now: by supposition this equation holds exactly for real gases. Yet an identification of temperature with a 'complex mean energy' is not possible for a different, more fundamental, reason. Consider the property denoted by 'complex mean energy'. The particles constituting the gas each have definite translational, vibrational, rotational and potential energy values (by supposition). Thus, the gas has a definite value for 'complex mean energy'. But the gas has a definite value for this property even when it is not in equilibrium. If temperature is identical to this property, it follows that the gas has a temperature outside of equilibrium too. However, this is in contradiction to the axiomatic laws of thermodynamics: *a gas does not have a temperature outside of equilibrium.* Temperature is that property which systems in equilibrium have in common! So this will not work either.

The SOSist is as ingenuous as she is persistent however. She will try to save the identification in some other way. One attempt is to *extend* the concept of temperature to what might be called 'non-equilibrium' temperature' and identify this with 'complex mean energy'. A second way is to attempt to identify temperature with 'complex mean energy' in equilibrium. I dub the latter '*local* local identification'. I consider these in turn below.

4.4.2.3 Extending the Concept of Temperature

The only definition of temperature is the one given by the axiomatic laws of thermodynamics and it is exactly this that rules out 'non-equilibrium temperature'. So the question is whether one can extend the concept of temperature to non-equilibrium. The point of such an extension is that it would then seem to be possible to identify temperature with the aforementioned 'complex mean

energy'.[34]

Writing about temperature Sklar suggests that indeed temperature has been extended. With reference to the relation between temperature and what he calls its 'statistical mechanical surrogates' he writes:

> "[T]he association of temperature with the measures of order and disorder in statistical mechanics leads to a *natural* extension of the absolute temperature concept in that theory... [For certain systems one can] describe the situation as one in which the system goes form a temperature of "minus zero" degrees, through finite negative temperatures that go "down" to "minus infinity"... Here, the concept extension follows in a *natural* way from the formalism designed to handle the more usual cases." (Sklar 1993 354 emphasis added)

Without an account of in what sense the concept extension is 'natural' this attempt to save identification is left wanting. But even granting that in some intuitive sense in which 'temperature' *can be* naturally extended to non-equilibrium, what would be the result of such an extension?

Here once again it is important to, as it were, tread carefully. Suppose that one extended the notion of temperature beyond equilibrium and identified temperature with 'complex mean energy'. This would be just to accept as true the equation relating temperature with 'complex mean energy'. Recall, however, that we do not have a precise mathematico-physical expression for this equation. In this sense, it is unclear exactly what one is identifying temperature with. But whatever the exact form of this expression, one is now guaranteed not to be picking out the same property as before viz. thermodynamic temperature: thermodynamic temperature just is that property which systems in thermodynamic equilibrium have in common. We can extend the notion of temperature, and take it to refer to a real property and identify this property with some statistical mechanical property. But this property, call it 'generalised-non-equilibrium-temperature', does not refer to the same property as 'temperature'

[34]What is puzzling about the question is the notion of possibility it involves. *What would it be to assert that it is possible to extend the concept, and what would it be to deny it?* In short, it is unclear on what grounds this question to be settled. Compare the situation with mathematics. Many mathematical concepts are extended. For example, the concept of number is extended to imaginary numbers. In mathematics, the possibility (or not) of extending a concept just turns on internal consistency. I contend that internal consistency is too weak for physics however. I will set this issue aside.

does (or better: the property that 'temperature' picks out at the possible world at which thermodynamics is literally true.) In having extended the notion of temperature we are now no longer talking about the very same notion we started with nor are we referring to the same property! In this case, and whatever else, the SOSist has not achieved ontological simplification as she set out to do.

4.4.2.4 *Local* Local Identification

A final attempt to save some form of ontological simplification is to proffer a '*local* local identification'. Taking the cue from Lewis (cf. section 4.4.2), the idea is to identify temperature with 'complex mean energy' for a specific system *in equilibrium*. It is 'local' twice over in the sense that one first considers only one kind of gas, say, (to avoid the problem detailed in section 4.4.2.3) and one 'localises' again by considering only equilibrium. The suggestion is that temperature be identified with 'complex mean energy' given this double localisation.

At first glance, the logic of the suggestion seems impeachable: by 'localising' in this way one avoids the problems that identification faces. But first glance is deceiving: the delineation that the putative identity requires, vis. 'in equilibrium', is itself untenable. Equilibrium is a thermodynamic property (in fact, as shown in section 4.2.1, equilibrium is a primitive property of thermodynamics) and, given the metaphysical 'picture' of kinetic theory, there is no such property.[35] That is: if a gas is just a collection of particles with various forms of energy, the properties of the gas - the various energies of the particles - are in constant flux. One can *define* equilibrium in kinetic theory: equilibrium is (sometimes) defined by the Maxwell-Boltzmann distribution for example. But the Maxwell-Boltzmann distribution is, strictly speaking, counterfactual, for it supposes that the particles only have pair-wise short-range interactions. Again, strictly speaking, there is no such thing as equilibrium in the thermodynamic sense for real gases[36] notion from the point of view of the kinetic theory (i.e. at PW_R). Thus, *local* local identity fails too.

[35] Better: If the kinetic theory is literally construed, then the metaphysical picture that this gives is such that there is no equilibrium.

[36] Recall from section 4.2.1 that equilibrium in the thermodynamic sense is strictly speaking an time-aysmmetric property of a system.

4.4.2.5 Temperature as an Emergent Property

Bridge-laws need not be identities. One argument for bridge-laws qua correlations of properties is that this is all that is justifiable from an evidential point of view. I touched on this at the start of section 4.4.1. Alternatively, one might argue for correlations by default: if bridge-laws qua identities fail then all bridge-law could be is an expression for the correlation (nomic or otherwise) of the properties. In the section proceeding this one I shall argue this is not the case either.

Here I consider an alternative view of temperature: this is the claim that temperature, far from being identifiable with a statistical mechanical property, is an emergent property. This is a claim put forward by Bishop and Atmanspacher (2006).

Bishop and Atmanspacher argue that the standard derivation of the Boyle-Charles law via kinetic theory

> "suggests a fairly straightforward [identification] of thermodynamic temperature [with mean kinetic energy]. Such a rough picture, however, would be a gross mischaracterization, based on a too generous treatment of some important details. Bishop and Atmanspacher (2006, 1769) [37]

We agree about *this* conclusion. But why do they think this is so, and what are the putative consequences? Here is the rest of the passage:

> "[T]he very concept of temperature is fundamentally foreign to statistical mechanics and has to be introduced, e.g., on the basis of phenomenological arguments. Thermal equilibrium is formulated by the zeroth law of thermodynamics: if two systems are both in thermal equilibrium with a third system, then they are said to be in thermal equilibrium with each other. (In this sense, the definition of temperature is relational.) Based on this equivalence relation, the phenomenological concept of temperature can be introduced in the usual text-book way. *Since thermal equilibrium is not defined at the level of statistical mechanics, temperature is not a mechanical property but,*

[37] In the original text it is written: "...straightforward *reduction* of thermodynamic temperature..." but their use of 'reduction' here is that of property identification (ontological reduction) and not intertheoretic reduction hence I have substituted terms accordingly.

rather, emerges as a novel property at the level of thermodynamics." (*ibid. emphasis added*)

Let me start with this part of the argument: Bishop and Atmanspacher are right to say that the concept of temperature is foreign to statistical mechanics (although it is unclear in what sense it is 'fundamentally' so). However, they are wrong in claiming that temperature is introduced on the 'basis of phenomenological arguments'. As we have seen, the bridge-law is a particular kind of theoretical stipulation. There is nothing *phenomenological* about it.

I think that Bishop and Atmanspacher are correct in arguing against the identity claim: temperature cannot be identical with a statistical mechanical property because temperature is that property which systems in thermal equilibrium have in common, and from the point of view of the statistical mechanics thermal equilibrium is, strictly speaking, a fiction. But their argument for temperature being an 'emergent' property is unpersuasive.

Bishop and Atmanspacher offer (something close to) a definition for what they call 'contextual emergence':

> "The description of properties at a particular level of description (including its laws) offers necessary but not sufficient conditions to derive the description of properties at a higher level. This version, which we propose calling contextual emergence, indicates that contingent contextual conditions are required in addition to the lower-level description for the rigorous derivation of higher-level properties." (Bishop and Atmanspacher, 2006, 1757)

Bishop and Atmanspacher's claim that temperature is an emergent property amounts to the claim that the description of statistical mechanical properties provide only necessary but not sufficient conditions for the derivation of a description of temperature. This, in turn, ushers in the ascription of 'emergent' to temperature.

It is unclear what it is for descriptions of properties at one level of description to provide both necessary and sufficient conditions to derive the descriptions of properties at a higher level. (Nor indeed, for them to provide only necessary but not sufficient; only sufficient but not necessary; or indeed neither necessary nor sufficient conditions.) This way of characterising the relation between the higher-

and lower-level properties is obscure - what does it even mean to say that descriptions offer conditions (of any stripe) for the derivation of other descriptions?! To be sure, temperature is not one of the properties of statistical mechanics nor do these properties - those of statistical mechanics, that is - alone provide necessary and sufficient conditions for the derivation of the higher-level law - the Boyle-Charles law - which features temperature.

Perhaps what is intended is that a property is emergent just in case the laws in which it features are not derivable from the laws of the lower-level theory alone. But this notion of 'emergence' is empty: in order to derive the laws of one theory from those of another, various auxiliary assumptions are invariably needed. But if that is right, all properties are 'emergent' in this sense and the ascription of 'emergent' to a property is uninformative.

4.4.3 Ontological Simplification: The Upward Path

The problem we have been grappling with rests on the assumption that 'temperature' denotes a real property in the world. Our job then seemed to be to work out which property that is. In particular, the philosopher's fixation has been to try to identify it with some statistical mechanical property - we have fixated on SOS in vain. But we cannot identify temperature with a single statistical mechanical property - there simply is no single statistical mechanical property with which the single property of temperature could be identified with.

The way out of this quagmire is to give up the assumption *qua assumption*. Let me explain. The point of reduction is to explain why a theory is as empirically successful as it is, based on the supposition that another theory, the reducing one, is true. If we have a successful NN reduction of the to-be-reduced theory to the reducing theory then we have good reason not to be ontologically committed to the to-be-reduced theory.[38] In short, this is because we have an explanation of why the reducing theory is (to the extent that it is) empirically successful *without* being literally true. We are not ontologically committed to the properties that the theory, literally construed, posits because the reasons which might ontologically committed us to the posits of a theory, viz. its empirically adequacy, are - in virtue of a successful reduction - redundant. We are ontologically committed to the posits of just those theories that are part of our best conceptual scheme,

[38]Recall, ontological simplification is not a necessary condition for NN reduction - it is a potential upshot.

which, *ceteris paribus* are the empirically most successful.[39] In case of successful reduction, we are not committed to the to-be-reduced theory, because, if you will, we do not need to be. We can account for empirical success of the theory by being committed to another theory, namely the reducing one. Or put another way: we have an explanation of why it is (at least to some extent) *as if* there are the properties posited by the to-be-reduced theory in the functional relationships given by its laws.

Reconsider the case of temperature. We have good reason to think that temperature is not a real property. (i.e. that 'temperature' fails to refer to a property.) Why? The reduction consists of the supposition that classical mechanics and the molecular hypothesis is literally true and in this 'metaphysical picture' there is not a single property that plays the role that temperature plays in the laws of thermodynamics. Indeed, the laws of thermodynamics are strictly speaking false in the possible world in which the classical mechanics is true. After all, we have to make counterfactual assumptions with respect to that world to derive them. As such, we search for an identity of the properties in vain. Yet, we have, *ex hypothesi*, an explanation of why such laws are empirical adequate to the extent that they are; we have an explanation of why it is (at least to a certain extent) as if there is a property like temperature which stands in a functional relationship to other thermodynamic properties as per the laws of that theory. The explanation of this empirical success is not undermined by the lack of cross-theoretic identities.[40]

4.5 Framework for Classical Statistical Mechanics

In this section I briefly set out the framework of classical mechanics, which is shared by both the Gibbsian and Boltzmannian schools.[41] Classical mechanics is most usefully presented in its Hamiltonian formulation in the context of statistical mechanics.

In its Hamiltonian formulation a system is described as consisting of particles,

[39] This is the essence of the Quinian meta-ontological position. (cf. chapter 1)

[40] Of course, if the reduction fails, then this might be good ground for being ontologically committed to thermodynamics, for, *ex hypothesi*, we would not have an explanation of the why there are empirically successful laws couched in terms of certain properties. That is, we would have reason to think that we ought to be ontologically committed to these properties for they are now contenders to be part of our best conceptual scheme, a lá Quine. cf. chapter 1.

[41] Here I follow [Frigg 2008].

which have definite positions, q and momenta p. The state of the system is determined by the positions and momenta of the particles. A system with N particles can be represented by a point, x, in a $6N$-dimensional phase space, Γ:

$$x := (q, p) := (q_1, \ldots, q_{3N}, p_1, \ldots, p_{3N}) \in \Gamma \tag{4.14}$$

The phase space, Γ, is endowed with a Lebesgue measure, μ. (This is also often just called the 'standard' or 'volume' measure.) The system evolution is determined by its Hamiltonian, $H(q, p, t)$:

$$\dot{p}_i = -\frac{\partial H}{\partial q_i} \text{ (for } i = 1, \ldots, 3N) \tag{4.15}$$

and

$$\dot{q}_i = \frac{\partial H}{\partial p_i} \text{ (for } i = 1, \ldots, 3N) \tag{4.16}$$

A trajectory through Γ represents the evolution of the system. Attention is restricted to the Hamiltonians for which the system is deterministic, in the following sense:

$$\forall x \in \Gamma : x \text{ lies on a unique trajectory through } \Gamma \tag{4.17}$$

i.e. no two trajectories in Γ cross. $H(q, p, t)$, thus, defines a one parameter group of transformations, ϕ_t, mapping Γ onto itself:

$$\forall (x \in \Gamma \ \& \ t) : x \to \phi_t(x) \tag{4.18}$$

This is usually called the 'phase flow'. Properties of the system are represented by functions, f of the form $f(q, p, t)$ (or just $f(q, p)$). The time evolution of such a function is given by the following:

$$\dot{f} = \{f, H\} + \frac{\partial f}{\partial t} \tag{4.19}$$

where $\{\ ,\ \}$ is the *Poisson bracket* given by:

$$\text{For all differential functions, } a \ \& \ b \text{ on } \Gamma : \{a, b\} := \sum_i \left[\frac{\partial a}{\partial q_i} \frac{\partial b}{\partial p_i} - \frac{\partial b}{\partial q_i} \frac{\partial a}{\partial p_i} \right] \tag{4.20}$$

If H does not explicitly depend on time, i.e. is a *conserved* quantity, then motion of the representative point, x is restricted to a $6N-1$ dimensional hypersurface of Γ, Γ_E, which is defined by $H(q,p) = E$, where E is the total energy of the system. This hypersurface is usually called the 'energy hypersurface'.

There is an important theorem pertaining to such Hamiltonian systems, *Liouville's Theorem*. Roughly, this states that the Lebesgue measure of a region of the phase-space is invariant under Hamiltonian follow. That is:

$$\text{For all Lebesgue measurable regions } R \subseteq \Gamma \text{ and } t : \mu(R) = \mu(\phi_t(R)) \tag{4.21}$$

Speaking loosely, in geometrical terms, whilst the *shape* of the region, R, may (and indeed generally will) change, the *volume* it occupies in the phase-space (i.e. its *phase volume*) remains constant. Liouville's Theorem also holds for the case in which $H(q,p) = E$, provided the measure on the energy hyper-surface, Γ_E is carefully chosen.

$$\text{For all } R_E \subseteq \Gamma_E \text{ and } t : \mu_E := \int_{R_E} \frac{d\sigma_E}{\parallel \text{grad}H \parallel} \tag{4.22}$$

where $d\sigma_E$ is a surface element on Γ_E, and $\parallel \text{grad}H \parallel$ is:

$$\parallel \text{grad}H \parallel := \left[\sum_{k=1}^{N} \left(\frac{\partial H}{\partial p_k} \right)^2 + \left(\frac{\partial H}{\partial q_k} \right)^2 \right]^{\frac{1}{2}} \tag{4.23}$$

Thus Liouville's Theorem in the case of $H(q,p) = E$ becomes:

$$\text{For all } R_E \subseteq \Gamma_E \text{ and } t : \mu_E(R) = \mu_E(\phi_t(R_E)) \tag{4.24}$$

4.6 Gibbsian Statistical Mechanics

In this section, I consider whether the Boyle-Charles law reduces to the classical Gibbsian statistical mechanics. That is, I apply the NN model of reduction to the derivation of the Boyle-Charles law from Gibbsian classical statistical mechanics. This will further illuminate the important aspects of the NN model and will pave the way for applying it to the case of the Second Law in section 4.7.

As said, I present a different derivation of the Boyle-Charles law. This falls under what physicists usually call statistical mechanics 'proper' - Gibbsian sta-

tistical mechanics. This is, as Frigg puts it, "the practitioner's workhorse" (Frigg 2008 178) However, calling it statistical mechanics 'proper' is at best an expression of sentiment: as Frigg shows (op. cit.), statistical mechanics has no canonical formulation and consists, rather, of a variety of different approaches which, if unified at all, are unified in positing the molecular hypothesis and involving some probabilistic reasoning. What is important to note is that Gibbsian statistical mechanics is far more general than the kinetic theory of gases. It can be used to derive far more than just the Boyle-Charles law, or other constitutive laws for gases. Indeed standard physics textbooks suggest that the entirety of thermodynamics can be derived in this manner. (cf., for example, Huang (1987, 127)) The hallmark of this approach is the use of *ensembles*.[42]

4.6.1 Re-deriving the Boyle-Charles Law

Consider a gas made up of N number of particles. As per the previous section, the state of the system is completely and uniquely specified by $6N$ canonical momenta and coordinates, $p_1, \ldots, p_{3N}; q_1, \ldots, q_{3N}$. One represents the state of the gas by a point in a $6N$-dimensional position-momentum space, Γ. The dynamics of the system are determined by the canonical equations of motion, equations 4.15 and 4.16 and the evolution of the system over time is represented by a trajectory through Γ.

Now consider some set of macroscopic constraints. There are many different *possible* states of the system which are compatible with these constraints. For example, consider a gas in a container of fixed volume: the particles of the gas are distributed in just one way however there are (infinitely) many ways the particles of the gas *can* be distributed within the container. Thus, whilst the system is only ever in one state, there are an infinite number of different states it *could be in* given the macroscopic constraints. This observation grounds the use of *ensembles*.

The centrepiece of the Gibbsian approach are *ensembles*.[43] An *ensemble* (for some system) is the (uncountably infinite) collection of *ensemble members*, where each *ensemble member* is an independent system governed by H, but distributed over different states. Thus, each *ensemble member* is represented by a point in

[42] Although the notion of ensembles was first introduced by Boltzmann.

[43] Here I follow the discussion in Huang (1987). Similar derivations can be found in other standard textbooks. For a more philosophically sensitive overview see Frigg (2008).

Γ.[44]

The *ensemble* is represented by a density function, a function of the momenta and positions of the particles over time:

$$\rho(p_1, \ldots, p_{3N}; q_1, \ldots, q_{3N}; t) \tag{4.25}$$

such that

$$\rho(p_1, \ldots, p_{3N}; q_1, \ldots, q_{3N}; t)\ d^{3N}p d^{3N}q \tag{4.26}$$

equals the relative density of *ensemble members* in the volume element $d^{3N}p d^{3N}q$ in Γ at time, t.[45] For what follows, it is useful to abbreviate equation 4.25 to

$$\rho(p, q, t) \tag{4.27}$$

In the Gibbsian context, Liouville's theorem yields the following[46]:

$$\frac{\partial \rho}{\partial t} + \sum_{i=0}^{3N} \Big(\frac{\partial \rho}{\partial q_i} \frac{\partial H}{\partial p_i} - \frac{\partial H}{\partial q_i} \frac{\partial \rho}{\partial p_i} \Big) = 0 \tag{4.28}$$

This means that ρ acts like an incompressible fluid on Γ. The further restriction that ρ does not depend explicitly on time, entails that:

$$\frac{\partial}{\partial t} \rho(p, q) = 0 \tag{4.29}$$

One now posits a fundamental assumption of the approach: *a priori equiprobability*. This assumption is that the actual state of the system is equally likely to be given by any of the *ensemble members*.[47] To model an equilibrium situation, one uses a particular kind of ensemble, namely the *microcanonical ensemble*[48] The *microcanonical ensemble* is defined by the following density function:

[44]Note that the *ensemble members* do not in anyway interact with each other; indeed they are just 'mental' or 'fictitious' copies of the actual system.

[45]That is, relative to all the ensemble members. It is important to note that this not the number of particles in $d^{3N}p\ d^{3N}q$.

[46]One proof is given by Huang (1987, 77).

[47]This is a controversial premise but I assume for the sake of the exposition. Notice that this is, in effect, an auxiliary assumption in the derivation. More about this shortly.

[48]There are other kinds of ensembles, viz. 'canonical' and 'grandcanonical' ensembles. These shall not concern us here.

$$\rho(p,q) = \begin{cases} \text{Constant} & \text{if } E < \mathcal{H}(p,q) < E + \Delta \\ 0 & \text{otherwise} \end{cases} \tag{4.30}$$

where $E < E + \Delta$ is an energy hypersurface *shell*, representing the system having an energy between $E < E + \Delta$.

The Gibbsian now proceeds to define the entropy of the system in terms of the microcanonical ensemble. One proceeds by first defining the volume of the phase space which the microcanonical ensemble occupies:

$$\Gamma(E) \equiv \int_{E<\mathcal{H}(p,q)<E+\Delta} d^{3N}p d^{3N}q \tag{4.31}$$

Similarly one introduces the following definitions

$$\Sigma(E) \equiv \int_{\mathcal{H}(p,q)<E} d^{3N}p d^{3N}q \tag{4.32}$$

and

$$\Sigma(E+\Delta) \equiv \int_{\mathcal{H}(p,q)<E+\Delta} d^{3N}p d^{3N}q \tag{4.33}$$

such that

$$\Gamma(E) \equiv \Sigma(E+\Delta) - \Sigma(E) \tag{4.34}$$

With the microcanoncial volume defined and this approximation introduced, one now defines the entropy of the system as follows:

$$S(E,V) \equiv k \log \Gamma(E) \tag{4.35}$$

where k is a constant.

One then justifies this definition by showing that it "possess all the properties of the entropy function in thermodynamics" (Huang, 1987, 13) In particular, that it is an extensive property (i.e. $S = S_1 + S_2$ for subsystems of the system, 1 and 2) and that it satisfies the second law of thermodynamics. Essentially, the justification is that for an isolated system in equilibrium, the entropy is a nondecreasing function of increase in volume (increase in volume being the only possible change for such a system) and thus the entropy of an isolated system cannot decrease as required. It can be shown that $S(E,V)$ is equivalent to $k \log \Sigma(E)$ "up to additive

constant terms of order $\log N$ or smaller" (Huang, 1987, 134). It is this form of the entropy that is used in the derivation of the Boyle-Charles law in terms of the microcanonical ensemble. The use of the microcanonical ensemble is *generic* in the sense that we have yet to specify any of the particulars of the system of interest, in the case in hand an ideal gas.

The Hamiltonian for an ideal gas is given by:

$$\mathcal{H} = \frac{1}{2m}\sum_{i=1}^{N} p_i^2 \tag{4.36}$$

where N is the number of particles for the system, each with a mass, m and a certain momentum denoted by p_i. As with the previous derivation, the Hamiltonian expresses the ideal gas assumption that the particles only have translational kinetic energy. Introducing a constant h to make $\Sigma(E)$ dimensionless, it can be shown that the Hamiltonian yields the following:

$$\Sigma(E) = C_{3N}\Big[\frac{V}{h^3}(2mE)^{3/2}\Big]^N \tag{4.37}$$

where

$$C_{3N} = \frac{\pi^{3N/2}}{(3N/2-1)!} \tag{4.38}$$

From $S(E,V) \equiv k\log\Sigma(E)$ we get:

$$S(E,V) = k\Big[\log C_{3N} + N\log\frac{V}{h^3} + \frac{3}{2}N\log 2mE)\Big] \tag{4.39}$$

Using the following limit,

$$C_n \xrightarrow{n\to\infty} \frac{n}{2}\log\pi - \frac{n}{2}\log\frac{n}{2} + \frac{n}{2} \tag{4.40}$$

it follows that

$$S(E,V) = k\Big[(\frac{3}{2}N\log\pi - \frac{3}{2}\log\frac{3}{2}N + \frac{3}{2}N) + N\log\frac{V}{h^3} + \frac{3}{2}N\log 2mE)\Big] \tag{4.41}$$

Some purely algebraic manipulations then show that the entropy $S(E,V)$ for the ideal gas system takes the form:

$$S(E,V) = Nk \log \left[V \Big(\frac{4\pi m}{3h^2} \frac{E}{N} \Big)^{3/2} \right] + \frac{3}{2} Nk \tag{4.42}$$

Solving equation 4.42 for E and noting that the total energy for the system is the internal energy U, it follows that:

$$U(S,V) = \Big(\frac{3}{4\pi} \frac{h^2}{m} \Big) \frac{N}{V^{2/3}} \exp \Big(\frac{2}{3} \frac{S}{Nk} - 1 \Big) \tag{4.43}$$

One then invokes the relation between temperature and entropy as per equation 4.44 below. Temperature is the partial derivative of the internal energy with respect to the entropy at fixed volume:

$$T = \Big(\frac{\partial U}{\partial S} \Big)_V \tag{4.44}$$

Substituting equation 4.43 into equation 4.44 and solving for temperature yields:

$$T = \frac{2}{3} \frac{U}{Nk} \tag{4.45}$$

Next, one invokes a similar equation for pressure. Pressure is negatively proportional to the partial derivative of the internal energy with respect to volume at fixed entropy:

$$P = -\Big(\frac{\partial U}{\partial V} \Big)_S \tag{4.46}$$

Similarly, substituting equation 4.43 into equation 4.46 and solving for pressure yields:

$$P = \frac{2}{3} \frac{U}{V} \tag{4.47}$$

Finally, substituting equation 4.45 into equation 4.47 and rearranging yields the Boyle-Charles law:

$$PV = NkT \tag{4.48}$$

This completes the derivation of the Boyle-Charles law via the microcanonical ensemble in the Gibbsian approach.

4.6.2 Gibbsian Statistical Mechanics: Reduction?

As presented, the above is literally the textbook derivation of the Boyle-Charles law using the microcanonical ensemble. What I wish to do in this section is to show how the Neo-Nagelian model of reduction provides a normative framework in which to think about this derivation.

The first thing to notice is that from a purely formal point of view this *is* a derivation of the Boyle-Charles law. On the Nagelian and NWR accounts, this would count as a *bona fide* reduction. But it is obvious that, when it comes to the question of whether this derivation *explains* the (empirical success of the) Boyle-Charles law, what matters is whether there is *warrant* for the various auxiliary assumptions and bridge-laws used in the derivation. It is patently absurd to have a model of reduction which puts no constraints on the auxiliary assumptions and bridge-laws. Yet, remarkably, every model hitherto put forward is entirely silent on this very point![49]

On problem for the Gibbsian apropos *warrant* for auxiliary assumptions is that of the limiting assumption involved in equation 4.40. The limit essentially takes the number of particles for the system to tend towards infinity. If the Gibbsian is going to argue for the reduction of TD to Gibbsian statistical mechanics, she must justify this assumption.

A broader and well-known problem with the Gibbsian approach is the recourse to ensembles.[50] This finds its clearest expression when one considers whether these is *warrant* for the bride-law. As with the derivation of the Boyle-Charles law from the kinetic theory of gases the requisite bridge-law is not made explicit. It is implicitly taken that one associates the thermodynamic and Gibbsian entropies. i.e. $S_{SM} \equiv S_{TD}$. Once the bridge-law has been introduced all the other thermodynamic properties can also be defined in terms of the statistical mechanical entropy: the bridge-law states the equivalence of the thermodynamic and statistical mechanic entropy allowing for a definition of the thermodynamic properties in terms of S_{SM} via the *thermodynamic functions* given in section 4.2. Schematically this can be represented as follows:

Start with the thermodynamic function for temperature:

[49] I would say that the most significant aspect of the Neo-Nagelian model of reduction is that it brings this issue to the fore.

[50] A rich discussion of the problems facing Gibbsian statistical mechanic reduction of thermodynamics is to be found in Callender (2001).

$$T \equiv \left(\frac{\partial U}{\partial S_{TD}} \right) \tag{4.49}$$

Then, using the bridge-law:

$$S_{SM} \equiv S_{TD} \tag{4.50}$$

one gets:

$$T \equiv \left(\frac{\partial U}{\partial S_{SM}} \right) \tag{4.51}$$

A similar result holds for pressure.[51]

Is this bridge-law *warranted*? We need to show that there is both *formal consistency* and *conceptual fit*. As we saw some justification is given by practitioner's, as per Huang (1987) above, for *formal consistency*. Specifically, it is shown, first, that like S_{TD}, S_{SM} is an extensive property, and, second, that in equilibrium, S_{SM} is a non-decreasing function of increase in volume, which in turn is the only permissible change to the system thus construed. However, the question of *conceptual fit* is tricky: S_{SM} may well be an extensive property but it is an extensive property of an ensemble.

Ensembles are infinite collections of copies of the actual system. This prompts the question of their status:

> "Are ensembles really an irreducible part of the Gibbsian scheme or are they just an expedient, or even a pedagogical ploy, of no fundamental significance? If so, how can the theory be reformulated without appeal to ensembles?" Frigg (2008, 143-144)

With regards the later, the answer is that it cannot: ensembles really are part and parcel of the Gibbsian approach and as such treating them as an expedient simply won't do. But what of a reduction based on them? An attitude one may have is to deny that they are problematic apropos reduction. After all, they *are* part and parcel of the reducing theory. Sure, one might not be satisfied with an explanation based on some such reducing theory (i.e. the explanation that a

[51] It is also important to note that the thermodynamic functions are inter-derivable *within statistical mechanics*. That is, starting with the entropy bridge-law and forming equation 4.51 one can derive the other thermodynamic functions as per section 4.2.

reduction affords is only as good as the reducing theory) but this is external to the reduction proffered *per se*. That is, it is a distinct question to whether thermodynamics reduces to Gibbsian statistical mechanics. However, this attitude belies the real problem here: the problem is not recourse to ensembles *simpliciter* but how one 'connects' ensembles to the the behaviour of real systems. That is, Gibbsian statistical mechanics can be explanatory so long there is an explanation supporting 'connection' between the dynamics of the ensemble and the dynamics of the system of interest.

If that is correct, then the focus of the discussion naturally moves on to the *a priori equi-probability* assumption. Recall, the assumption is that the actual state of the system is equally likely to be any one of the possible states that make up the ensemble. I contend that the conceptual problems besetting the Gibbsian approach is precisely the problem of showing that this assumption is explanation supporting. Indeed getting *conceptual fit* is a well-known (albeit not by this name) and outstanding problem with the Gibbsian account. (cf. Callender (1999), for example.) My aim here is not try to settle this problem - I am merely pointing out that the question of reduction is dependent on whether this problem can be solved.[52]

To summarise: there is a derivation of Boyle-Charles law from Gibbsian statistical mechanics. The NN model of reduction acts as a normative framework to assess whether this constitutes a reduction. To my mind, the biggest problem for the Gibbsian account is justifying the recourse to ensembles, and this remains an open problem.[53] In the next section, I consider another derivation of a thermodynamic law, namely the derivation of the Second Law from Boltzmannian statistical mechanics.

[52]There is on-going research into the relation between Gibbsian and Boltzmannian statistical mechanics. (cf., e.g. Lavis 2008). In section 4.6 I argue that *conceptual fit* for the Boltzmannian bridge-law looks promising. If there is a tight-link between the approach, it may be argued that the *conceptual fit* in the latter carries over to the former. However, this too remains an open question.

[53]Note again this is somewhat loose talk: it is theories which reduce not laws. What I mean here is that, were each of the laws of thermodynamics derived in a similar way from Gibbsian statistical mechanics, this would not constitute a reduction of thermodynamics to Gibbsian statistical mechanics.

4.7 Boltzmannian Statistical Mechanics

4.7.1 Reducing the 2nd Law

In this final section I consider the derivation of a different thermodynamic law, the Second Law. Deriving the Second Law of thermodynamics is a well-known problem in statistical mechanics and various attempts at doing so have been made. There is no consensus about which, if any, of these attempts is correct; indeed each of the putative solutions to it have problematic features. (For an overview see Frigg (2008).)

For the coming discussion, it is necessary to make a distinction between two versions of the Second Law for the coming discussion: the *Strict Static* Second Law and the *Strict* Second Law.

The *Strict Static* Second Law (Second Law_{SS}) is the law given in section 4.2.1. It is an universal law in the sense that it putatively holds for all systems, not just, say, gases or a particular kind of liquid.[54] What is important to notice is that the law is silent about the approach to equilibrium and an increase in entropy. It requires only that the entropy of an isolated system does not decrease, where the entropy of system is just that property which is maximized at equilibrium. There is no mention of an approach to equilibrium at all.

The *Strict* Second Law (Second Law_{S}) takes the approach to equilibrium into account: the entropy of a system monotonically increase towards a local maximum, and once in this equilibrium state, the system remains in it. Like the Second Law_{SS}, it is also a universal law.

There is a general consensus, at least amongst Boltzmannians, that attempting to derive the Second Law_{S} is misguided.[55] (cf. Callender (2001); Sklar (1993); Frigg (2008)).) Interestingly there are two different reasons given for this. One pertains to motivation: attempting to derive the Second Law_{S} is misguided because there is a lack of motivation to do so. In this context Callender is usually cited: attempting to derive the Second Law_{S} would be to take "thermodynamics too seriously." (cf. Callender (2001).) After all, real systems do not show strict monotonic increase in entropy nor do they remain in the equilibrium state exactly - the entropy increases, *roughly* monotonically, and then fluctuates around the

[54] In fact one ought to be more careful: as it is one of the axiomatic laws, if it does not hold for a particular system then that system is not a *Thermodynamic System.* (cf. section 4.2.1)

[55] On the Gibbsian approach, one can derive the Second Law_{SS} but one has a problem then accounting for Second Law_{S}. cf. Frigg (2008, 140-173).

equilibrium value. Frigg echoes the point about motivation but also cites another reason, namely expedience:

> "[T]he Second Law cannot be derived from SM [statistical mechanics]. The time reversal invariance of the dynamics and Poincaré recurrence imply that the Boltzmann entropy does not increase monotonically at all times. In fact, when an SM system has reached equilibrium it fluctuates away from equilibrium every now and then." (Frigg, 2008, 139)[56]

That the Second Law$_S$ cannot be derived from statistical mechanics is widely claimed. (cf. Frigg (2008) for further references.) If it cannot be derived then it is misguided to try, of course.

Both the claim that there is a lack of motivation to derive Second Law$_S$ and that it is expedient not to try need further examination.

As regards the lack of motivation, one has to be clear in distinguishing reducing the Second Law$_S$ to SM and explaining why the relevant systems behave in the way that they do by SM. If one's aim is to explain why there is a general tendency for systems to tend towards equilibrium (albeit not strictly monotonically and with fluctuations out of equilibrium) then focusing on deriving the Second Law$_S$ misses the point. In this sense, Callender's point stands. However, if one is interested in the reduction of the thermodynamics to statistical mechanics, deriving the Second Law$_S$, irrespectively of it being strictly speaking false, is what is at stake. Thus, from the point of view of reduction, there is *motivation* to try to derive the Second Law$_S$. Recall that the aim of reduction is to provide an explanation of the extent of the empirical success of a theory (or particular law of a theory) from the reducing theory. Second Law$_S$ is not literally true but it is a robust generalization, which statistical mechanics aims to explain.

What about the claim that Second Law$_S$ 'cannot be derived' from statistical mechanics? As is the case with reductions in general, any putative derivation of the Second Law$_S$ involves more than just the reducing theory (in this case, statistical mechanics) but also auxiliary assumptions and bridge-laws.[57] The claim that Second Law$_S$ cannot be derived needs qualification, for presumably

[56] Assume for the moment that there is a bridge-law to the effect that the Boltzmann entropy, S_B, is directly proportional to the thermodynamic entropy, S_{TD}. This is needed to 'connect' Boltzmann's Law to thermodynamics; it presupposed in this passage. I return to this below.

[57] As per the previous footnote, I return to the entropy bridge-law shortly.

one *can* add *some* dynamical auxiliary assumption to get the relevant asymmetry and thus derive the law. But I think that we have an inuition that some such derivation is not right: deriving the law from some such auxiliary assumption (whatever its exact form) undermines the sense in which we are reducing the Second Law$_S$ to *statistical mechanics*. Some such auxiliary assumption is, we intuit more than a *modification* of the dynamics of the reducing theory.

An attractive feature of the Neo-Nagelian model is that it frames this intuition. It is best seen by looking at a concrete case which exemplifies this kind of case. Consider the research program referred to as 'stochastic dynamics'. Roughly characterised, on this approach one coarse-grains the phase-space of the system and postulates a probabilistic dynamics which determines transitions across the cells of the coarse-grained partition. (cf. Uffink (2007, 1038-1063)) Cast into the Neo-Nagelian model, we have here a Dynamical auxiliary assumption. Suppose for the sake of argument that via this auxiliary assumption one can exactly derive Second Law$_S$. The important issue from the point of view of NN reduction, as I have been at pains to stress throughout this thesis, it that there needs to be *warrant* for auxiliary assumptions.[58] Is there *warrant* for such an auxiliary assumption? That is, does it pass the test that in making it less counterfactual one derives more empirically adequate laws? (cf. chapter 1.4.5.) It does not (indeed cannot) for some such auxiliary assumption cannot be made less counterfactual with respect to the reducing theory in the requisite way. The 'probabilistic' system that the auxiliary assumption renders does not lay on a continuum to the kind of Hamiltonian systems that statistical mechanics postulates. The time-aysmmetric dynamics given by the stocastic dynamics cannot be made less counterfactual with respect to the reducing theory for time-aysmmetry is an all-or-nothing affair. Thus, the test as to whether or not one gets an empirically more adequate law in making the assumption less counterfactual is a *non-starter* and, therefore, some such auxiliary assumption would lack *warrant*. Indeed, any auxiliary assumption which introduces the time-asymmetry encoded into thermodynamics and in particular the Second Law, is not going to be a *warranted* auxiliary assumption.

The right way, then, to think about the original claim is not that that one cannot *derive* Second Law$_S$ from Boltzmannian statistical mechanics, but that one cannot *reduce* it to Boltzmannian statistical mechanics, for the auxiliary

[58]There needs to be *warrant* for the bridge-laws too but here we are concerned with the auxiliary assumptions.

assumptions used in the derivation will not be *warranted*.

How then to proceed? The leading suggestion in the Boltzmannian camp is that one derives 'thermodynamic-like' behaviour in lieu of the Second Law$_S$: the idea is to show that systems tend, albeit not monotonically, towards equilibrium and once there tend to stay there albeit on average and with small fluctuations. Or expressed in terms of the Boltzmann entropy, what one wants to show is that systems' Boltzmann entropy increase, albeit non-monotonically towards the local maximum and fluctuates around that equilibrium value.[59]

Suppose that one has a bridge-law associating the Boltzmann entropy, S_B with the thermodynamic entropy, S_{TD} and that one can show that for any given system its Boltzmann-entropy exhibits 'thermodynamic-like' (TD-like) behaviour. Would we then have reduced the Second Law$_S$ to Boltzmannian statistical mechanics? Schaffner's modification of Nagel's model is usually cited to argue that we would have, for deriving the exact law of the to-be-reduced theory is not a necessary condition for reduction. (cf. Frigg *op. cit.* and references therein.)[60] On the Neo-Nagelian account too, one need not derive the exact laws of the to-be-reduced theory, and so one need not derive Second Law$_S$. (cf. chapter 1.4) But what the Neo-Nagelian model stresses, which the Schaffner-modified Nagelian model misses entirely, is that *merely deriving* TD-like behaviour is not enough: we need to show that there is *warrant* for the bridge-law and the auxiliary assumptions used in the derivation. If one can derive TD-like behaviour from *warranted* auxiliary assumption and bridge-law, then, *ipso facto*, on the Neo-Nagelian account one will have reduced Second Law$_S$ to Boltzmannian statistical mechanics.[61]

Significant inroads have been made in this respect. In what follows I consider very recent work due to Frigg and Werndl (forthcoming). Frigg & Werndl rehabilitate the Ergodic Program, proffering a derivation of TD-like behaviour based on, so-called, 'epsilon-ergodicity'. Whilst Frigg & Werndl are not successful in

[59] Admittedly there is a certain ambiguity to what constitutes 'thermodynamic-like' behaviour. I return to this issue below - bracket it for the moment.

[60] Of course, I argued, in chapter 1.3.2, that the Schaffner-modified Nagelian model is not tenable but here I am just reporting that Schaffner's modification of Nagel's model is usually appropriated in this context.

[61] It is important to stress that what I am referring to in broad terms as the Boltzmannian approach is actually one of several different approaches each of which use the Boltzmannian framework set out in the next section, section 4.7.2. A discussion of the other approaches which fall under the Boltzmannian umbrella can be found in, for example, Uffink (2007) and Frigg (2008).

deriving TD-like behaviour for all kinds of thermodynamic systems they do derive it for gases, and, moreover, there is reason to be optimistic about the prospects of extending this derivation to liquids and possibly even gases.

I proceed as follows. In section 4.7.2 I set up the Boltzmannian Framework. In section 4.7.3 I introduce the Ergodic Program and the notion of epsilon-ergodicity and show how this allows one to derive Boltzmann's Law. In section 4.7.4, I review how Frigg & Werndl deal with the putative problems that the Ergodic Problem faces and finally in section 4.7.5 I reflect on some limitations of the approach and future prospects.

4.7.2 The Boltzmannian Framework

In the Boltzmannian framework one considers a isolated system comprised of n particles with three degrees of freedom, occupying a finite volume V such that the system's total energy is fixed at E.[62] The state of the system is represented by a point, x - the system's fine-grained micro-state - in its $6n$ dimensional phase space Γ_γ, and the system's dynamics is governed by its Hamiltonian, H. Given the volume and energy constraints, x is restricted to a finite sub-region of Γ_γ: a $6n-1$ dimensional hypersurface, $\Gamma_{_E}$, known as the 'energy hypersurface'. The phase space is endowed with a Lebesgue measure $\mu_{_L}$, which induces a measure $\mu_{_E}$ on the energy hypersurface. Given this measure, one can speak of the 'volume' (or 'hypervolume') of subsets of Γ_γ and Γ_E.

As set out in section 4.5, the Hamiltonian defines a measure preserving flow ϕ_t on Γ_γ, meaning that $\phi_t : \Gamma_\gamma \rightarrow \Gamma_\gamma$ is a one-to-one mapping and $\mu_{_L}(R) = \mu_{_L}(\phi_t(R))$ for all times t and all regions $R \subseteq \Gamma_\gamma$, from which it follows that $\mu_{_E}(R_E) = \mu_{_E}(\phi_t(R_E))$ for all regions $R_E \subseteq \Gamma_E$. Intuitively, this determines all the possible trajectories of x on the energy hypersurface.

Let M_i, $i = 1, ..., m$ represent the system's macro-states. These are characterised by the values of macroscopic variables such as local pressure, local temperature, and volume. A conceptual cornerstone of the Boltzmannian approach is that a system's macro-state supervenes on its fine-grained micro-state: for every fine-grained micro-state $x \in \Gamma_E$ there corresponds *exactly one* macro-state. However, different micro-states can determine the same macro-state; the relation between the micro-states and macro-states is many-to-one.[63] Γ_{M_i} form a

[62] Here I am following the exposition given by Frigg (2008). This takes it cue from Lebowitz (1993, 1999) and Goldstein and Lebowitz (2004).

[63] Frigg makes a distinction between macro-states and macro regions - $\Gamma_{M_i} := \{x \in \Gamma_E \,|\, M_i =$

partition of Γ_{E}.[64]

One then defines the Boltzmann entropy for a macrostate:

$$S_{B}(M_i) \;=\; k_{B}\;\log[\mu_{L,\,E}(\Gamma_{M_i})], \tag{4.52}$$

where k_{B} is Boltzmann's constant. We are interested in the system's entropy at time, t. This is defined as follows:

$$S_{B}(t) := S_{B}[M(x(t))]. \tag{4.53}$$

where $M(x(t))$ is the system's macro-state at time t. The equilibrium state of the system is defined as just that macro-state in which the Boltzmann entropy of the system is maximal - call this M_{Eq}. For convenience, denote all non-equilibrium macro-states by $M_{\neg Eq,j}$, where $j = 1, ..., m-1$. (This is just to say that there are $m-1$ non-equilbirbrium macro-states, the number one gets by taking all the macro-states and excluding the equilbirum macro-state.)

The aim is now to show that systems exhibit TD-like behaviour, characterised in terms of the Boltzmann entropy.

How can one derive TD-like behaviour? One traditional answer is given within, the so-called, 'Ergodic Program': if a system is ergodic, then TD-like behaviour follows.[65] The Ergodic Program, discussed shortly, has been widely criticised and is taken by many to be a dead-end. However, Frigg and Werndl (forthcoming) argue that these criticisms can be met once one makes the transition to what is called 'epsilon-ergodicity'. In the next section I consider Frigg & Werndl's rehabilitation of the Ergodic program.

$M(x)\}, i = 1, ..., m,$, the subset of Γ_{E} consisting of all fine-grained micro-states that correspond to macro-state M_i. As Frigg says: "The proposition that a system with energy E is in macro-state M_i and the proposition that the system's fine-grained micro-state lies within Γ_{M_i} always have the same truth value," (Frigg 2008 104) although in some contexts it is important to distinguish between them. For the present purpose this distinction is not needed.

[64] To be more precise Γ_{M_i} form a partition of $\Gamma_{\gamma,\,a}$, where $\Gamma_{\gamma,\,a}$ is the accessible region of the phase-space, which lies within Γ_{E} but this detail does not matter for our purposes.

[65] It has been forcefully argued by Frigg (2008); Frigg and Werndl (forthcoming) that any derivation which does not make recourse to dynamics is untenable. In particular Frigg (2008); Frigg and Werndl (forthcoming) argue that so-called 'typicality approaches' are untenable precisely because on these accounts do not make any recourse to systems' dynamics.

4.7.3 The Ergodic Program

Ergodic theory was developed in the context of dynamical systems theory.[66] The general system we considered above can be construed as a dynamical system, in the following sense: $(\Gamma_{_E}, \mu_{_E}, \phi_t)$ is a dynamical system where $\Gamma_{_E}$ is (the accessible part of) the energy hypersurface of the system's phase-space as before, $\mu_{_E}$ is the standard Lebesgue measure on the phase-space but renormalised to one on $\Gamma_{_E}$, and ϕ_t is measure-preserving flow on the phase-space as per the system's Hamiltonian.

Consider the phase-flow, $\phi_t(x)$ on $\Gamma_{_E}$. Now consider a solution starting at an arbitrary point $x \in \Gamma_{_E}$ and a measurable set $A \subseteq \Gamma_{_E}$.[67] The time-average for this solution with respect to A is:

$$L_A(x) = \lim_{t\to\infty} (1/t) \int_0^t \chi_A(\phi_t(x))dt. \tag{4.54}$$

where χ_A is the characteristic function of A. Birkhoff's pointwise ergodic theorem proves that $L_A(x)$ exists for all $x \in \Gamma_{_E}$ except for at most a set B of measure zero, i.e. at most there is a set $B \subseteq \Gamma_{_E}$ such that $\mu_E(B) = 0$

$(\Gamma_{_E}, \mu_{_E}, \phi_t)$ is *ergodic* if and only if for any measurable set A:

$$L_A(x) = \mu_E(A) \tag{4.55}$$

for any $x \in \Gamma_{_E}$ except at most a set of measure zero. Before discussing the significance of this theorem in relation to Botlzmann's Law, it is useful to characterise the notion of epsilon-ergodicity too.

In order to define epsilon-ergodicity, we first generalise the notion of ergodicity to the notion of ϵ-ergodicity.[68] ϵ-ergodicity is defined as follows. A dynamical system of the above kind, $(\Gamma_{_E}, \mu_{_E}, \phi_t)$, is ϵ-ergodic iff:

> there is a set $Z \subset \Gamma_{_E}$ with $\mu(Z) = \epsilon$, where $\epsilon \in \mathbb{R}$ and $0 \leq \epsilon < 1$; and with $\phi_t(\hat{\Gamma}_{_E}) \subseteq \hat{\Gamma}_{_E}$ for all t, where $\hat{\Gamma}_{_E} := \Gamma_{_E} \backslash Z$, such that $(\hat{\Gamma}_{_E}, \mu_{\hat{\Gamma}_{_E}}, \phi_t^{\hat{\Gamma}_{_E}})$ is ergodic, where $\mu_{\hat{\Gamma}_{_E}}(.) := \mu_{_E}/\mu_{_E}(\hat{\Gamma}_{_E})$ for any measureable set in $\hat{\Gamma}_{_E}$ and $\phi_t^{\hat{\Gamma}_{_E}}$ is ϕ_t restricted to $\hat{\Gamma}_{_E}$.

[66] Here again I follow (Frigg and Werndl, forthcoming).

[67] The solution is the unique trajectory through the point x.

[68] This is phonetically hazardous terminology: 'the-Greek-letter'-ergodicity is the generalise notion of ergodicity and the name of said Greek letter in English is used to denote the kind of ergodicity on which Frigg & Werndl base their account.

An ergodic system is the special case of an ϵ-ergodic system where ϵ is set to equal 0. We now define an epsilon-ergodic system:

> A dynamical system, $(\Gamma_{_E}, \mu_{_E}, \phi_t)$, is epsilon-ergodic iff there exists a very small ϵ (i.e. $\epsilon << 1$) for which the system is ϵ-ergodic.

Thus, an epsilon-ergodic system is one which is ergodic on the vast majority of the energy hypersurface, $\Gamma_{_E}$.

The consequences of ergodicity are powerful.[69] Coarsely put, if a system is ergodic then an arbitrary trajectory through $\Gamma_{_E}$ is such that the amount of time that trajectory spends in a particular macro-state M_i is equal to the size of that macro-state $\mu_E(M_i)$ on $\Gamma_{_E}$.

How does all this relate to TD-like behaviour? For gases, M_{Eq} is vastly larger than any other $M_{\neg Eq}$; $\Gamma_{_E}$ is almost entirely taken up by it.[70] This combined with assumption that the system is ergodic allows one to derive TD-like behaviour. Here is how Frigg & Werndl put it:

> The dynamics will carry x to $[M_{Eq}]$ and will keep it there most of the time. The system will move out of the equilibrium region every now and then and visit non-equilibrium states. Yet since these are small compared to $[M_{Eq}]$, it will only spend a small fraction of time there. Hence the entropy is close to its maximum most of the time and fluctuates away from it only occasionally. (Frigg and Werndl, forthcoming, 6)

It is tempting to try to make this statement more precise by making recourse to probabilities. Thus TD-like behaviour could be characterised by the conjunction of the following three conditionals. 1) if the system is in a particular non-equilibrium state $M_{\neg Eq,1}$ it is likely to evolve either into M_{Eq} or another non-equilibrium state $M_{\neg Eq,2}$, such that $S_{_B}(M_{\neg Eq,2}) \geq S_{_B}(M_{\neg Eq,1})$. 2) If a system is in M_{Eq} it is likely to remain in that state. 3) Moreover, if the unlikely does occur and it moves out of M_{Eq} and into an arbitrary $M_{\neg Eq,j}$, then 1. Whilst there is much intuitive appeal to this idea, making it rigorous is not forthcoming for two reasons: first, one cannot unproblematically define probabilities on the account. At best, one can regard the time-averages associated with different macro-sates as

[69] I return to epsilon-ergodicity shortly.

[70] cf., for example, Goldstein (2001).

a proxy for the probability of the system being in that macro-state. But, second, even if one does so, the 'probabilities' that this yields cannot be construed as conditional ones (as the above characterisation, from 1 - 3, would have it). That a system is ergodic does not entail anything about its likely behaviour conditional on its current state. Rather, ergodicity ('only') entails that time-averages associated with different macro-states equal the phase-volumes of those macro-states: given the relative sizes of the phase-volumes of the macro-states, this entails just the characterisation of TD-like behaviour proffered by Frigg & Werndl above.

A different kind of probability does enter into the picture, however. If a system is ergodic it behaves TD-like with probability 1. If a system is epsilon-ergodic, it behaves TD-like with probability $1 - \epsilon$. Notice that these are claims about the probability that a system which is (epsilon-) ergodic behaves TD-like and not a claim about the entropy profile of the system, i.e. it does not tell us about the likely size of fluctuations of the entropy. These probabilities are a consequence of the ergodic theorem, as per equation 4.55: phase-averages and time-averages are equal except for at most a set of measure zero. What this means is that there are some points in Γ_{E} for which the system fails to be ergodic but that these points are of measure zero. In case of epsilon-ergodicity such points are of measure $1-\epsilon$.

4.7.4 Solving Problems with the Ergodic Program

The 'Classic' Ergodic Problem sketched out above faces two well-known problems: the so-called *Measure Zero* problem and, following Frigg & Werndl, what I will call the *Irrelevancy* problem.

4.7.4.1 The *Measure Zero* Problem

The *Measure Zero* problem: the equality of the time-average with the phase-average as per equation 4.55 holds for all $x \in \Gamma_{E}$ except for, at most, a set of measure zero, B.[71] The problem is that it might be the case that a system starts on $x_B \in B$ and if so, the relevant solution (i.e. the trajectory through x_B) does not exhibit TD-like behaviour. It is also often pointed out B can be, intuitively speaking, rather large despite being of measure zero with respect to Γ_{E} (cf. Sklar 1993 182-188 for example.)

Frigg & Werndl's response to the *Measure Zero* problem that it is only a problem if one wants to derive the Second Law$_S$.

[71] Notice that an epsilon-ergodic system the size of the measure of the set of 'bad' points is ϵ.

> "[The *Measure Zero*] criticism is driven by the demand to justify a strict version of the second law [i.e. Second Law$_S$], but this is, as argued in the last section, an impossible goal. [cf. section 4.7.1.] The best one can expect is an argument that TD-like behaviour is very likely, and the fact that those initial conditions that lie on non-TD-like solutions have measure zero does not undermine that goal. Consequently, we deny that the measure zero problem poses a threat to an explanation of TD-like behaviour in terms of ergodicity [and indeed to epsilon-ergodicity]." (Frigg and Werndl, forthcoming, 7)

One must be careful in reading Frigg & Werndl's response to the *Measure Zero* problem. When they state that the "criticism is driven by the demand to justify a strict version of the second law" (*ibid.*), their notion of 'strict' is ambiguous between 'strict' as monotonic increase in entropy and no fluctuations out of equilibrium (i.e. the sense in which Second Law$_S$ is strict as per the above) and 'strict' as applying to all possible evolutions of every system (i.e. the sense in which Second Law$_S$ is universal). In discussion with Frigg, he has emphasised to me that the intended meaning of 'strict' here is the latter, viz. universality. Of course the two are not mutually exclusive and Frigg was quick to point out that neither strict (in my sense) nor universal thermodynamic behaviours are forthcoming. However, Frigg & Werndl's claim, as per the above quote, is that failing to show that *all* trajectories are TD-like (i.e. the failure of the universality) does not undermine the putative explanation; showing that TD-like behaviour is very likely (specifically, with probability 1 given ergodicity and $1-\epsilon$ given epsilon-ergodicity) is explanatorily sufficient. However, Frigg & Werndl's claim that a derivation of a 'strict' version of the Second Law is 'an impossible goal' (*op. cit.*) is not quite right, indeed precisely because it is ambiguous between the two sense of 'strict' just discussed. Using my senses of 'strict' and 'universal', the right thing to say is that it is impossible to derive the strict Second Law[72] - i.e. one can only derive TD-like behaviour - but it has not been shown that is impossible to derive TD-like behaviour universally. The crucial point is that the impossibility of the former does not entail the impossibility of the latter.

Be that as it may, Frigg & Werndl are correct to bite the bullet on the issue: showing that TD-like behaviour is very likely is explanatorily sufficient. That

[72]Understood as the claim that any derivation of it would involve auxiliary assumptions for which there is no *warrant*. cf. section 4.7.1.

is, showing that TD-like behaviour is very likely is sufficient, *ceteris paribus*, for reduction.

However, we can also make the bullet less bitter by making two points. Oddly they do not find articulation in the relevant literature. First, in case of ergodicity, the set of 'bad' initial conditions is *at most* of measure zero. But that does not imply that there are such 'bad' initial conditions; it just means that one cannot show that there are none. The logic of this point carries over the case of epsilon-ergodicity too: the set of 'bad' initial conditions are *at most* of measure ϵ. Second, it is not clear that the 'bad' initial conditions are actually *bad*: there may be some initial conditions that do not lie on TD-like solutions, however, just what the evolution of the system is like on such initial conditions is unknown. Importantly, it is not demonstrated that they are 'anti-TD-like', where 'anti-TD-like' means, roughly, that the system evolves from a high-entropy macro-state into an extremely low one and fluctuates around *that* state. Thus, the *Measure Zero* is not a knock-down argument against the (epsilon-) ergodicity approach.

4.7.4.2 The *Irrelevancy* Problem

I now turn to the *Irrelevancy* problem. The *Irrelevancy* problem: the actual systems with which SM is concerned are not ergodic so ergodicity is irrelevant to Boltzmann's Law. Specifically, there are two widely discussed theorems which putatively show that the relevant systems are not ergodic: the Kolmogorov-Arnold-Moser theorem (KAM theorem) and the Markus-Meyer theorem (MM theorem). If that is correct then whatever follows from ergodicity is irrelevant to the behaviour of SM systems.

Before considering this problem further, it is important again to emphasize that there is a difference between how successful a reduction of the to-be-reduced theory to the reducing theory is, and how empirically adequate the laws of the to-be-reduced theory are. A successful NN reduction requires that the (albeit not necessarily exact) laws of the to-be-reduced theory be derived from the reducing theory, and auxiliary assumptions and bridge-laws for which there is *warrant*. The auxiliary assumptions need not be true to be *warranted*. In general, the *warrant* for the auxiliary assumptions is not determined by their counterfactualness - the auxiliary assumptions need to be counterfactual to afford the derivation. For, from the point of view of the reducing theory, the to-be-reduced theory is, again in general, strictly speaking, false.

With this distinction in mind, isn't the *Irrelevancy* problem irrelevant? The aforementioned theorems putatively show that real systems are not ergodic but ergodicity is an auxiliary assumption in the derivation of TD-like behaviour and it *need not be true* to be *warranted*. As per the previous paragraph, the counterfactualness, *per se*, of an auxiliary assumption does not undermine its *warrant*. Recall the derivation of the Boyle-Charles law: one assumed various counterfactual auxiliary assumptions in this derivation. The test for the *warrant* for the auxiliary assumptions was not their counterfactualness but that making them less counterfactual yielded an empirically more adequate law, namely Van der Waals equation. (cf. section 4.3.2 above.)

The present case is different in one very important respect. Unlike the Boyle-Charles law, which is strictly speaking false, TD-like behaviour is true![73] Systems *do* behave TD-like. Given this, a derivation of TD-like behaviour from counterfactual auxiliary assumptions cannot constitute an explanation of it, and, *ipso facto*, some such derivation cannot constitute a reduction of the Second Law$_S$. It precisely because TD-like behaviour - unlike Second Law$_S$ - is taken to be a true generalization of the behaviour of real systems that the KAM and Markus-Meyer theorems, which putatively show that the relevant systems are not ergodic, are thought to be so damaging to the Ergodic Program.

Frigg & Werndl argue that neither the KAM nor the MM theorems show that the relevant systems are not, at least, epsilon-ergodic, as they are taken to. The technicalities need not concern us here; I will just summarise their results.

The KAM theorem pertains to integrable Hamiltonian systems which are subjected to a small non-integrable perturbations. The energy hypersurface of an integrable Hamiltonian system is foliated into tori, on each of which there is periodic or quasi-periodic motion. What the theorem states is that when such a system slightly perturbed in the above sense, some of the tori survive the perturbation whilst others do not with the following consequence:

> "The region on Γ_E in which the tori survive and the region in which they break up are both invariant under the dynamics. The motion on the region with surviving tori cannot be ergodic (or epsilon-ergodic)

[73] It might be argued that TD-like behaviour as characterised does not capture the whole truth about the behaviour of real systems but that is a different issue. If one's logic is binary, and one agrees that TD-like behaviour is not false (and surely it is not false, for nothing contradicts it!), then it is true! One might also worry that it is not specific enough apropose size of fluctuations. I consider this issue briefly below.

> because the solutions are confined to tori. Therefore, dynamical systems to which the KAM-theorem applies are not ergodic, and for a small enough perturbation, they are not epsilon-ergodic either..." (Frigg & Werndl 2011 7)

Thus, the KAM theorem is usually taken to show that many (if not all) SM systems are not ergodic. (cf. for example, Earman and Rédei (1996, 70) and Sklar (1993, 172).) If that's right, then from the point of NN reduction, the ergodic auxiliary assumption used in the derivation of Boltzmann's Law cannot be *warranted* as discussed above.

Frigg & Werndl point out an important and overlooked aspect of the theorem. The theorem only applies to 'extremely small' perturbations; for larger perturbations the aforementioned 'surviving' tori also disappear. Here is a quote from Pettini to this effect:

> [F]or large n-systems[74] - which are dealt with in statistical mechanics - the admissible perturbation amplitudes for the KAM-theorem to apply drop down to exceedingly tiny values of no physical meaning" (Pettini 2007 60 (as cited in (Frigg & Werndl 2011 7))

Given that the 'surviving' tori also disappear for larger perturbations, the motion on Γ_E can be (epsilon-) ergodic, if not fully ergodic. Another way to think about this: it is only for extremely small perturbations that Γ_E is foliated in such a way, to use Earman and Rédei's phrase, as "to contain islands of stability where the flow is non-ergodic" (Earman & Rédei 1996 70) - any larger and these islands are thought to disappear too. The latter claim is not strictly proven but Frigg & Werndl argue that, *in so far as SM systems can be construed as integrable Hamiltonian systems that are subjected perturbations at all*, there are good theoretical reasons to think that this is the case for systems with many degrees of freedom. (cf. Frigg & Werndl 2011 11) These theoretical considerations are supported by numerical investigations too:

> "For perturbations higher than a specific moderate perturbation, nearly all or all of the energy hypersurface seems to be taken up by irregular motion, and hence the motion appears to be epsilon-ergodic...

[74] Here, 'n' refers to the systems number of degrees of freedom.

> It might be that very small islands of regular motion persist for arbitrary large perturbations. But then these regular regions are very small, and so while the system would fail to be ergodic, it will still be epsilon-ergodic. Furthermore, there is evidence that, everything else being equal, the main region of ergodic behaviour grows larger and larger as the number of degrees of freedom increases..." (Frigg & Werndl 2011 12)

Thus, the KAM-theorem does not show that the relevant systems with which SM is concerned are not ergodic and moreover there are good reasons to think that they are, at least, epsilon-ergodic.

Next Frigg & Werndl consider the MM theorem. Their strategy is similar to the previous one: the theorem is shown to be, in the sense given below, inapplicable to the relevant systems of interest. Again, the technicalities need not detain us; I shall just present the main results.

The MM theorem is a topological theorem about the function space, Λ, of all infinitely differentiable Hamiltonians on a compact space, M. To understand the theorem it is necessary to introduce the notion of an *(epsilon-) ergodic Hamiltonian*.

> An *(epsilon-) ergodic Hamiltonian*, H_ϵ, is one which has a dense set of energy values for which the flow on the energy hypersurface is (epsilon-) ergodic.

The MM-theorem states that the set of (epsilon-) ergodic Hamiltonians in Λ is meagre.[75] Why is this problematic for the derivation of TD-like behaviour? Here is how Frigg & Werndl put it:

> "It is a plausible demand that physical properties be robust under small structural perturbations. [In the present case] this amounts to requiring that if a system is (epsilon-) ergodic, a system with a very similar potential function should be (epsilon-) ergodic as well." (Frigg & Werndl 2011 14)

[75] A set is *meagre* iff it is the countable union of nowhere dense sets. A set is *comeagre* iff its complement is meagre. Intuitively, *meagreness* is the topological analogue to the measure-theoretic notion of a set of measure zero. cf. (Frigg & Werndl 2011 14)

This is ruled out by the MM theorem.[76] However, Frigg & Werndl go one to point out that for relevant systems the MM theorem does not apply. In particular, the MM theorem pertains only to those Hamiltonians on compact spaces but "nearly all systems considered in classical mechanics have non-compact phase spaces" (Frigg & Werndl 2011 13). They also point out that appeal to the fact that Γ_E is typically compact has no purchase because, whilst that is true, the theorem is "about the *full* phase space Γ of a system and cannot be rephrased as a theorem about energy hypersurfaces." (*ibid.* orig. emph.)

Finally, Frigg & Werndl argue that the previous point notwithstanding the theorem is also inapplicable to the relevant systems in a different sense. The MM theorem is proved by showing that for generic Hamiltonians in Λ there is *exactly one* minimal value of the energy for which it is not the case that there is (epsilon-) ergodicity arbitrarily close to it. The salient point is this, according to Frigg & Werndl: such energy minima are unrealistic for, at the very least, SM gases.

Having, so to speak, warded-off the worries about the KAM and MM theorems, Frigg & Werndl go on to argue that theoretical and numerical considerations support the conjecture that SM gases are indeed epsilon-ergodic and as such the derivation TD-like behaviour, at least for gases, is successful.

If all this is right then to have a NN reduction of the Second Law$_S$ one only needs to further show that there is *warrant* for the bridge-law associating Boltzmann's entropy, S_B, with the thermodynamics entropy, S_{TD} for gases. This requires showing *formal consistency* and *conceptual fit*. On both accounts, things look promising. As regards the former, S_B is an extensive property of the system and obeys the Second Law (as per and in the above sense). This then matches the desiderata set out by Huang (op. cit.) in section 4.6.1. Moreover, one can show that S_B is equal to the S_{TD} for ideal gases up to an arbitrary additive constant. (cf. Frigg & Werndl 2010 15).

As regards *conceptual fit*, things are much better than in the Gibbsian case. S_B is a property of an individual system (and not an ensemble) and, crucially, the changes in a system's entropy are determined by its dynamics, rather than constructed out of an average of all its possible states. Thus the S_B-S_{TD} bridge-law is, *warranted* (or at least far more *warranted* than its Gibbsian counterpart).

[76]The technical formulation of the requirement is that for any $H_\epsilon \in \Lambda$ there is an open set in Λ around H_ϵ such that all the Hamiltonians in the open set are also (epsilon-) ergodic. The MM theorem rules this out.

4.7.5 Prospects and Limitations

TD-like behaviour, as discussed in section 4.7.1, is true. However, it is also not very specific: although it does say that large fluctuations cannot happen frequently and have to be short lived, it is desirable to get a sharper probabilistic bounds on the size of possible entropy fluctuations. Also it does not speak to relaxation times, the time it takes for systems to reach equilibrium.

Getting bounds on the size of fluctuations and getting relaxation times right remains an outstanding problem in the context of Boltzmannian statistical mechanics. With respect to relaxation times, it is worth noting that these are vastly different for different systems (which in itself implies that an attempt at a *general* characterisation of them would be misguided) and importantly that not all systems are such that they 'relax' quickly. (Frigg & Werndl (*ibid.*) give the example of cooling iron, which may take very many months to reach equilibrium!) As Frigg & Werndl note, epsilon-ergodicity is silent on the latter issue and, in fact, there are almost no analytic results. (cf. Frigg & Werndl 2011 18.)

4.8 Chapter Summary

In this chapter, I have substantiated the Neo-Nagelian model of reduction. I then went on to rebut the dominant view in the literature on reduction, namely that temperature is identical to mean kinetic energy. It is misguided to conceive of bridge-laws as in any way metaphysically substantive claims. Bridge-laws are a particular kind of theoretical stipulation.

In the second half of this chapter I applied the Neo-Nagelian model of reduction. I considered the Gibbsian derivation of the Boyle-Charles law and showed that it does not constitute a reduction. I then considered recent work rehabilitating the Ergodic program. The latter does constitute at least a partial reduction of the Second Law of thermodynamics and makes for a promising avenue for future research.

In the next, and final, chapter, I apply the Neo-Nagelian model of reduction to the derivation of the Second law of thermodynamics from quantum statistical mechanics.

Chapter 5

Neo-Nagelian Reduction and QSM

5.1 Chapter 5 Introduction

Quantum statistical mechanics, as the term is used here, is a branch of statistical physics based on quantum mechanics. Classical statistical mechanics (and the kinetic theory of gases) aims to account for the macroscopic behaviour of systems in terms of classical mechanics and various probabilistic assumptions. Likewise, quantum statistical mechanics aims to account for the macroscopic behaviour of systems in terms of quantum mechanics and probabilistic assumptions.[1]

There are three recent research projects in QSM which I consider here. The first is due to David Albert and is based on the GRW interpretation of quantum mechanics. I shall call this 'Albertian QSM'. The second is due to Hemmo and Shenker. They take their cue from Albert's approach but in place of Albert's recourse to GRWian quantum mechanics use decoherence models. I call this 'H&Sian QSM'. Finally there is the approach based on very recent work by Linden, Popescu, Short and Winter. I call this 'Aharonovian QSM'.[2] It is the latter with which I shall be primarily concerned.

Before saying more about each of these, let me first say what they have in common. In broad strokes, each is concerned with explaining the Second Law of

[1] It is important to note that QSM, in the sense intended here, has nothing to do with the statistics pertaining to quantum 'gases' as characterised by Bose-Einstein and/or Fermi-Dirac statistics.

[2] The framework takes it cue from certain works of Yakir Aharonov. cf. Popescu et al. (2006).

thermodynamics. Each of these approaches cites the ubiquity of Second-Law-like macroscopic phenomena and the need to explain it. Indeed Hemmo and Shenker explicitly motivate their approach in this respect:

> "Can we explain the laws of thermodynamics, in particular the irreversible increase of entropy...? Attempts based on classical dynamics have all failed." (Hemmo and Shenker, 2001, 555)[3]

Thus, each of these approaches is a reductionist enterprise, in the Neo-Nagelian sense. My aim in this chapter is to use the Neo-Nagelian model of reduction to assess whether they are successful in reducing the Second Law$_S$ to quantum statistical mechanics. As with chapter 4.7.1, the strategy is to derive TD-like behaviour. If TD-like behaviour is derived in the requisite Neo-Nagelian way, then, *ipso facto*, one will have reduced Second Law$_S$ to quantum statistical mechanics.

In the first section, section 5.2, I shall give a brief sketch of Albertian and H&Sian QSM and indicate their limitations. In section 5.3. I shall present Aharonovian QSM, as per Linden, Popescu, Short, and Winter (2009). In section 5.4 I shall then consider whether Aharonovian QSM affords a successful NN reduction of the Second Law$_S$. Finally in section 5.5 I consider the question of interventionism.

5.2 Albertian and H&Sian QSM

David Albert (Albert, 1994, 2000) advocates an approach to QSM based on *GRWian quantum mechanics* (cf. Ghirardi et al. (1986)). Indeed, Albert regards GRWian quantum mechanics as providing the only tenable resolution to the quantum measurement problem but an assessment of this claim is beyond the scope of this work.[4] The concern here is whether Albertian QSM gives an account of the approach to equilibrium.

Hemmo and Shenker (2001) provide a clear presentation of Albertian QSM and I give a simplified version based on theirs, sufficient for the current purpose. Consider some system comprised of n particles. Some such system is represented by a wavefunction; consider it expanded in the position basis. Under GRWian

[3] Hemmo & Shenker may well be too pessimistic in their appraisal of CSM (cf. chapter 4.7) but the point about explanation is the salient one at present.

[4] It is worth noting, however, that there is certainly nothing like a consensus on the issue of the measurement problem.

quantum mechanics the system evolves according to the standard Schrödinger dynamics except that it randomly localises sometimes. This localisation is the GRWian 'collapse' or 'jump'. These jumps are similar in effect to 'orthodox' collapse only that, rather than collapsing into an exact eigenstate of position, it localises around one. Note that there are two fundamental probabilities involved in GRW: the probability of 'jumps' occurring and the probability that the wave-function localises around a given eigenstate of the position operator. (The former is a new fundamental metaphysical posit; the latter is the analog to the standard Born Rule.) Hemmo and Shenker summarize the theory usefully:

> "It is convenient to think about the GRW dynamics of the system as inducing random perturbations of its Schrödinger trajectory. In a sense, the GRW jumps may be taken to alter in a random way the Schrödinger trajectory the system would follow had the jumps not occurred. This means that the GRW trajectory can be thought of as a patchwork of segments of different Schrödinger trajectories each of which corresponds to a different initial state of the system. The net result of the GRW dynamics is that the trajectory of the system is genuinely and irreducibly stochastic with the probabilities given by the GRW constants ... and the usual quantum mechanical Born rule." Hemmo and Shenker (2001, 558)

Given this, it is possible that a jump does not occur at all for a given system but the chance of this diminishes the larger the system. Call those trajectories in which jumps actually do occur, as per the relative frequencies suggest by the theory, 'GRW-normal' trajectories and those in which the jumps do not occur, or occur not in keeping with the relative frequencies, 'GRW-abnormal' trajectories. Albert is only concerned with the former, i.e. it is assumed that there are only 'GRW-ian normal' trajectories. Now consider two further kinds of trajectories (i.e. two subsets of all the 'GRW-ian normal' trajectories) that the system may follow: those for which the system tends towards equilibrium (i.e. entropy increasing ones) or those for which the systems tends away from equilibrium (i.e. entropy decreasing ones) and call them 'thermodynamic-normal' and 'thermodynamic-abnormal' respectively. With this framework in place, Albert's proposes the following conjecture:

> *Albert's Conjecture*: for any large quantum system there are vastly many

more thermodynamic-normal trajectories than thermodynamic-abnormal ones.[5]

Assuming that the system's *actual* trajectory is GRW-normal, it can be thought of as a patchwork of segments of different thermodynamic (either 'normal' or 'abnormal') trajectories. Given Albert's conjecture, the system will likely tend to towards equilibrium, for even if it starts out on a thermodynamic-abnormal trajectory it is likely to 'jump' to a thermodynamic-normal one, and even if it 'jumps' to an abnormal one from a normal one, it will likely not remain on it.[6]

However, the conjecture is just that, *a conjecture*: there is no proof that there are vastly more thermodynamic-normal trajectories than abnormal ones. In fact, there is not even a characterisation of thermodynamic trajectories on which a proof could be based.[7] Thus, whatever else one thinks about Albert's GRW approach, as it stands, it is only a *sketch* of an account of systems tending towards equilibrium.

Hemmo and Shenker (2001, 2003) also propose an approach to QSM but instead of making recourse to GRWian quantum mechanics, they utilise environmental decoherence models[8] associated with no-collapse theories. Their motivation is to provide an alternative to Albertian QSM:

[5]Actually Albert does not formulate this conjecture explicitly; rather such a conjecture is tacitly assumed. cf. Albert (2000, 151).

[6]Note that the system is not less likely to 'jump' when it is on a thermodynamic-normal trajectory – the chance of a 'jump' is the same irrespectively of which kind of trajectory the system is on. Rather, it is likely to 'jump' onto another thermodynamic-normal trajectory and this because there are many more of these than thermodynamic-abnormal ones. It is also interesting to note that this is a structurally similar argument to the case of ergodicity: thermodynamic-normal trajectories play an analogous role to 'ergodic-trajectories' in the classical Boltzmannian framework. (Recall 'ergodic-trajectories' are just those that exhibit TD-like behaviour and this kind of behaviour is the proxy for thermodynamic-like behaviour in that context.) The relative 'numbers' of the two kinds trajectories are, as per Albert's conjecture, intended to be similar to the relative 'numbers' of 'ergodic-trajectories' to 'non-ergodic-trajectories'. However, the difference is that on Albert's picture, *even if* a system starts out on a thermodynamic-abnormal one, it will, likely, 'jump' into a thermodynamic-normal one.

[7]Albert does provide some plausibility for the conjecture however. cf. Albert (2000, 155). It is also worth noting the following: Albert suggests that the GRW approach is the only approach to QSM which does not need to make recourse to two distinct kinds of probabilities and, presumably on grounds of parsimony, this is a reason to prefer it to the other approaches. "[T]he business of underwriting the thermodynamic regularities of the world, on any of the proposals for making sense of quantum mechanics I know of, with the sole exception (of course) of the GRW theory... is going to call for something along the lines of a *probability-distribution* over *initial wave-functions*, a probability-distribution which (note) is altogether *unrelated* and *in addition* to the probabilities with which those theories underwrite the statistical regularities of *quantum mechanics*." (Albert, 2000, 154, orig. emph.) I do not assess this argument.

[8]cf., for example, Joos and Zeh (1985) and Zurek et al. (1993).

> "If [Albertian QSM] were the only way to explain thermodynamics on the basis of quantum dynamics, the GRW approach would gain a serious advantage over its alternatives." (Hemmo and Shenker, 2001, 556)

Using the same framework as above, vis. thermodynamic-normal and -abnormal trajectories, they propose the following conjecture:

> *H&S Conjecture*: "perturbations of the molecules in the decoherence interaction are enough to put them with high probability on thermodynamic-normal trajectories." (Hemmo and Shenker, 2001, 556)

As said, their motivation is to show that Albert's GRW approach is not unique in giving an account of the systems tending towards equilibrium and in this they are successful, for their account rests on a conjecture no stronger than that required by Albert. For the very same reason, however, their's is not a fully-fledged account of systems tending towards equilibrium either.

A recent paper by Linden, Popescu, Short, and Winter (Linden et al., 2009) purportedly goes beyond conjecture:

> "We prove, with virtually full generality, that reaching equilibrium is a universal property of quantum systems." (Linden et al., 2009, 4)

Aharonovian QSM is certainly similar in spirit to Hemmo and Shenker's in that it is an interventionist account[9], and unlike Albert's, one which does not utilise 'collapse' in its results.

As I said right at the start of this chapter, Aharonovian QSM, like Albertian and H&Sian QSM, is a putative explanation of thermodynamic-like behaviour. The Neo-Nagelian model provides a normative and useful framework for critically engaging with putative explanations of this kind. If the derivation of TD-like behaviour, viz. the approach to equilibrium and fluctuations around that state, satisfies the Neo-Nagelian criteria, namely that the auxiliary assumptions and bridge-laws are *warranted*, then the derivation will be a reduction *qua* explanation of Second Law$_S$ to QSM.

[9]Hemmo and Shenker identify their account as interventionist (cf. Hemmo and Shenker (2001, fn. 11)) but do not discuss this aspect of it. See my chapter 5.5.

In section 5.3, I present Linden et al.'s argument that reaching equilibrium is a universal property of quantum systems. I present it without philosophical scrutiny.

In section 5.4, I will cast Linden et al's argument into the Neo-Nagelian mould. In particular, I consider exactly what the theorems allow one to derive, and the auxiliary assumptions and bridge-law used in this derivation. I then consider whether there is any *warrant* for the auxiliary assumptions and bridge-law. In so doing I diagnose a conceptual confusion in Linden et al's argument. However, I show that one can rehabilitate Aharonovian QSM in this respect, and, whilst there remain some important open questions, Aharonovian QSM does constitute at least a partial reduction of the Second Law. This section constitutes the most substantial part of the chapter.

In the final section of the chapter, section 5.5, I consider a potential problem with the conceptual framework of Aharonovian QSM in a broad sense. The Aharanov approach is an interventionist one but interventionism is considered to be a dubious stance in the context of the foundations of statistical mechanics. To make the Aharonovian QSM tenable one needs to show that interventionism does not undermine the *warrant* for the auxiliary assumptions and bridge-law.

5.3 Aharonovian QSM

How does the Aharonovian QSM purport to account for a particular system tending towards and remaining in equilibrium? One starts by considering a larger system - call this the 'global system'. The particular system of interest, i.e. the system which is to be shown to tend towards equilibrium, is a smaller system *within* the global system. Call the system of interest the 'subsystem'. The rest of the global system, i.e. just that part of the global system that is not the subsystem, forms the environment of the subsystem - call this the 'bath'. This is the (albeit somewhat abstract) 'metaphysical picture' of the Aharanov approach.

All the possible states of the global system are represented by a Hilbert space, $\mathcal{H}$. This is decomposed into the subsystem and the bath: $\mathcal{H} = \mathcal{H}_{\mathcal{S}} \otimes \mathcal{H}_{\mathcal{B}}$, where $\mathcal{H}_{\mathcal{S}}$ and $\mathcal{H}_{\mathcal{B}}$ are the Hilbert spaces of the subsystem and the bath, respectively. One upshot of the Aharanov approach is that nothings hangs on the particular features of the decomposition: the *formal* results hold for any arbitrary decomposition of

$\mathcal{H}$.[10] However, a physical assumption is that the subsystem is significantly smaller than the global system and, therefore, than the bath.

The global system, and, therefore, the subsystem and the bath, are posited to evolve according to a Hamiltonian, H. The Hamiltonian is "completely general" (Linden et al., 2009, 4) except for one constraint, namely that it is assumed to have non-degenerate energy gaps. What does this mean? Linden et al put it like this:

> "[A] Hamiltonian has non-degenerate energy gaps if [and only if] any nonzero difference of eigenenergies determines the two energy values involved." (Linden et al., 2009, 2)

Explicitly: for any four eigenstates with eigenvalues $E1$, $E2$, $E3$ and $E4$: if $E1-E2 = E3-E4$, then either ($E1 = E2$ & $E3 = E4$) or ($E1 = E3$ & $E2 = E4$). (cf. (Linden et al., 2009, 2)) More intuitively one can express this as follows: the difference between any two energy eigenvalues is unique.

I return to this assumption in section 5.4.3. The main strategy of Aharonovian QSM is to show that subsystems reach equilibrium in virtue of being embedded in global systems of some such dynamics.

To state and assess the central results of their paper, it is necessary to introduce some further technical machinery. (Here I follow their paper closely.) Let $\mid \Psi(t)\rangle$ denote the pure state of the global system at time, t. This state can be represented as a density matrix $\rho(t) = \mid \Psi(t)\rangle\langle\Psi(t) \mid$. The density matrices $\rho_{\mathcal{S}}(t)$ and $\rho_{\mathcal{B}}(t)$ represent the state of the subsystem and the bath, respectively. They relate to one another by tracing over the global state in the following way: $\rho_{\mathcal{S}}(t) = \mathrm{Tr}_{\mathcal{B}}\ \rho(t)$ and $\rho_{\mathcal{B}}(t) = \mathrm{Tr}_{\mathcal{S}}\ \rho(t)$.

Central to Aharonovian QSM are time-averages, *the time-averaged state of the system, target system and bath.* These are denoted by ω, $\omega_{\mathcal{S}}$, and $\omega_{\mathcal{B}}$ respectively and are defined as follows:

$$\omega = \langle\rho(t)\rangle_t = \lim_{\tau\to\infty}\int_0^\tau \rho(t)\ dt \tag{5.1}$$

$$\omega_{\mathcal{S}} = \langle\rho_S(t)\rangle_t = \lim_{\tau\to\infty}\int_0^\tau \mathrm{Tr}_{\mathcal{B}}\ \rho(t)dt \tag{5.2}$$

[10] At least part of the account for the ubiquity of equilibration is just this feature.

$$\omega_{\mathcal{B}} = \langle \rho_B(t) \rangle_t = \lim_{\tau \to \infty} \int_0^{\tau} \mathrm{Tr}_{\mathcal{S}}\ \rho(t) dt \tag{5.3}$$

Two more items are needed. First, the "effective dimension", d^{eff} of a mixed state, ρ

$$d^{\text{eff}}(\rho) = \frac{1}{Tr(\rho^2)} \tag{5.4}$$

Here is how Linden et al. characterise $d^{\text{eff}}(\rho)$:

> "This tells us, in a certain sense, how many pure states contribute appreciably to the mixture. In particular a mixture of n orthogonal states with equal probability has effective dimension n. Unlike the support of the density matrix, this notion captures the probabilistic weight of different states in the mixture, and is continuous." (Linden et al., 2009, 3)

Second, the trace-distance, D between two states:

$$D(\rho_1, \rho_2) = \frac{1}{2} Tr \sqrt{\rho_1 - \rho_2} \tag{5.5}$$

Again, in this section I just explicate their position:

> "The trace-distance characterises how hard it is to distinguish two states experimentally (even given perfect measurement). When it is small, the two states are indistinguishable. More precisely, it is equal to the maximum difference in probability for any outcome of any measurement performed on the two states." (*ibid.*)

Consider $D(\rho_{\mathcal{S}}(t), \omega_{\mathcal{S}})$ - the trace-distance between the state of the subsystem at a particular time and the time-averaged state of the subsystem. And now consider the time-average of *this* distance: $\langle D(\rho_{\mathcal{S}}(t), \omega_{\mathcal{S}}) \rangle_t$. With this machinery one can state the key theorem of their paper:

Theorem 1. *For any state* $\mid \Psi(t) \rangle \in \mathcal{H}$ *evolving under a Hamiltonian,* H *with non-degenerate energy gaps,*

$$\langle D(\rho_{\mathcal{S}}(t), \omega_{\mathcal{S}}) \rangle_t \leq \frac{1}{2} \sqrt{\frac{d_{\mathcal{S}}}{d^{\mathit{eff}}(\omega)_{\mathcal{B}}}} \leq \frac{1}{2} \sqrt{\frac{d_{\mathcal{S}}^2}{d^{\mathit{eff}}(\omega)}}$$

(Linden et al., 2009, 4), where $d_{\mathcal{S}}$ is the dimension of the Hilbert space for the entire system.

What does this theorem tell us? Here is how Linden et al. characterise it:

> "By bounding $\langle D(\rho_{\mathcal{S}}(t), \omega_{\mathcal{S}})\rangle_t$, [the] theorem tells us that the subsystem will equilibrate whenever the effective dimension explored by the bath $d^{\text{eff}}(\omega)_{\mathcal{B}}$ is much larger than the subsystem $d_{\mathcal{S}}$, or whenever the effective dimension explored by the total state $d^{\text{eff}}(\omega)$ is much larger than two copies of the subsystem... In other words, if $\langle D(\rho_{\mathcal{S}}(t), \omega_{\mathcal{S}})\rangle_t$ is small, the system spends most of its time close to the equilibrium state." (*ibid.*)

If this is persuasive then a further argument is needed to the effect that $\langle D(\rho_{\mathcal{S}}(t), \omega_{\mathcal{S}})\rangle_t$ *is* small. Linden et al provide a second theorem to this end:

Theorem 2. *i) The average effective dimension* $\langle d^{eff}(\omega)\rangle_\Psi$ *where the average is computed over uniformly random pure states* $|\ \Psi(t)\rangle \in \mathcal{H}_R \subset \mathcal{H}$ *is such that*

$$\langle d^{eff}(\omega)\rangle_\Psi \geq \tfrac{d_R}{2}$$

ii) For a random state $|\ \Psi(t)\rangle \in \mathcal{H}_R \subset \mathcal{H}$*, the probability* $\mathrm{Pr}_\Psi\{\mathrm{d}^{eff}(\omega) < \frac{\mathrm{d_R}}{4}\}$ *that* d^{eff} *is smaller than* $\frac{\mathrm{d_R}}{4}$ *is exponentially small, namely*

$$\mathrm{Pr}_\Psi\left\{\mathrm{d}^{eff}(\omega) < \tfrac{\mathrm{d_R}}{4}\right\} \leq 2\,\exp\left(-\mathrm{c}\sqrt{\mathrm{d_R}}\right),\ \textit{where}\ c \approx 10^{-4}$$

(Linden et al., 2009, 5)

Here is how Linden et al characterise the content of the second theorem:

> "Point (i) essentially tells us that the average effective dimension is larger than half the dimension of the Hilbert subspace [$\mathcal{H}_R$], so when we draw states [at random] from a subspace of large dimension, the effective dimension $d^{\text{eff}}(\omega)$ of a typical state is large. Point (ii) makes the result even sharper, telling us that the probability of having a small effective dimension is exponentially small." (Linden et al., 2009, 5)

Their idea is that, given that the dimension of the subsystem is small by hypothesis, it follows from Theorems 1 and 2, that $\langle D(\rho_{\mathcal{S}}(t), \omega_{\mathcal{S}})\rangle_t$ is typically small. Thus, they have purportedly shown that arbitrary small subsystems of large quantum systems typically tend towards and remain in equilibrium. Again, I ought to stress that here, I have merely presented the central tenets of their paper, without, so to speak, philosophical scrutiny.

5.4 NN Reduction and Aharonovian QSM

In this section, I cast Aharonovian QSM into the Neo-Nagelian mould. Having done so, I consider exactly what one can derive from the theorems, and the auxiliary assumptions and bridge-law needed.

Recall the overall structure of Linden et al.'s argument. (For convenience, I abbreviate '$\langle D(\rho_{\mathcal{S}}(t), \omega_{\mathcal{S}})\rangle_t$' to '$\langle D_{\rho,\omega}\rangle_t$'.) First, they state that:

> "[w]hen [$\langle D_{\rho,\omega}\rangle_t$] is small, the subsystem must spend almost all of its time very close to $\omega_{\mathcal{S}}$. In other words, when $\langle D_{\rho,\omega}\rangle_t$ is small, the subsystem equilibrates to $\omega_{\mathcal{S}}$." (Linden et al., 2009, 5)

By Theorem 1 $\langle D_{\rho,\omega}\rangle_t$ is bound below the ratio of twice the dimension of the target system, $d^2_{\mathcal{S}}$, to the effective dimension of the whole/global system, $d^{\text{eff}}(\omega)$. When the dimension of the target system is far smaller than the global system, i.e. when $d^{\text{eff}}(\omega)$ is far larger than $d_{\mathcal{S}}$, $\langle D_{\rho,\omega}\rangle_t$ must be very small.

By hypothesis the subsystem is small relative to the system as a whole, which is represented by $d_{\mathcal{S}}$ being small.[11] What needs to be shown, then, is that $d^{\text{eff}}(\omega)$ is large. The second theorem is intended to show this: it bounds $\langle d^{\text{eff}}(\omega)\rangle_\Psi$ above half the dimension of the global system (which is large by hypothesis). From this, suggest Linden et al, it follows that $\langle D_{\rho,\omega}\rangle_t$ is typically small. This, so their argument goes, shows that a small subsystem of a large global system will typically tend towards, and then remains, in, equilibrium.[12]

As it stands, this is not persuasive for a number of reasons. First, there is no argument that $\langle D_{\rho,\omega}\rangle_t$ being small is an adequate representation of the subsystem

[11] Linden et al assume that $d_{\mathcal{S}}$ being small does indeed represent the subsystem being small relative to the total system *without argument*. However, whether this is true is not at all obvious. The Aharonovian would need to argue for this assumption; this remains an open problem for the approach.

[12] Notice that this result holds for an arbitrary initial state of the subsystem. In particular, whether it is initially far from equilibrium does not impinge on the result.

reaching and remaining at equilibrium. This, as will be shown, tacitly involves the bridge-law in this context. I take this up in section 5.4.1 below. Second, it is not clear, I suggest, how the results in the second theorem relate to $\langle D_{\rho,\omega} \rangle_t$. In particular, it is not clear how, if at all, Theorem 2 shows that $\langle D_{\rho,\omega} \rangle_t$ is typically small. In section 5.4.2, I address this issue. In having addressed these issues, and making explicit what the auxiliary assumptions are, I offer a derivation of the desired conclusion at the end of section 5.4.2. With this in place, I then consider whether the auxiliary assumptions and bridge-law are *warranted*, in section 5.4.3.

5.4.1 Representing Equilibrium

In their paper, Linden et al take $\langle D_{\rho,\omega} \rangle_t$ being small to be a representation of the subsystem reaching and remaining at equilibrium. However they do not motivate this in any way! In this subsection I address several points to fill this conceptual gap and indicate their limitations.

The state of the global system, subsystem and bath are represented by density matrices and their evolution governed by a certain Hamiltonian. The time-averaged state of the subsystem, $\omega_{\mathcal{S}}$, is obtained by integrating the density matrix representing the subsystem with respect to time and evaluating it in the time-limit, $\lim_{\tau \to \infty}$. This is a standard procedure, resulting in an uncontroversial mathematical object. But what does this represent physically? Whilst it is not explicitly stated in the paper, $\omega_{\mathcal{S}}$ is tacitly assumed to represent the equilibrium state. Indeed, it is only by taking this to represent the equilibrium state that $\langle D_{\rho,\omega} \rangle_t$ relates to equilibrium at all. This is the bridge-law, in the nomenclature of Neo-Nagelian reduction: the thermodynamic equilibrium state is associated with the time-averaged state of the subsystem. Using the notation introduced for thermodynamics in chapter 4.2.1, the Aharonovian QSM bridge-law is:

$$\rho_{\mathcal{S}}(E_{\mathcal{S}}) = \omega_{\mathcal{S}} \tag{5.6}$$

I return to the question of whether there is *warrant* for this bridge-law below, in section 5.4.3.

With the equilibrium state defined, the trace-distance is then used to relate the state of the system at a particular time, $\rho_{\mathcal{S}}(t)$, to the equilibrium state, $\omega_{\mathcal{S}}$. What does the trace-distance represent? Formally the trace-distance "is equal to the maximum difference in probability for any outcome of any measurement performed on the two states." (Linden et al., 2009, 5) The issue is how to un-

derstand this physically. In particular, how to substantiate the claim that "[it] characterises how hard it is to distinguish two states" (*ibid.*).

First note that the formal claim is probabilistic. Conceptually, $D_{\rho,\rho\prime} = 0$ is not equivalent to the claim that the measurements performed on the two states will have the same outcome. Rather, $D_{\rho,\rho\prime} = 0$ says that the probability of getting different outcomes is zero.[13] The smaller the value of $D_{\rho,\rho\prime}$, the closer are the expectation values for any operator acting on each state; the more likely it is that the outcome of measurements performed on the two states will be the same.[14] Why is the probability of there being a difference in outcomes of measurements performed on states a suitable proxy for the physical (dis)similarity, or (in)distinguishability, of the states? The brief answer is that, owing to superpositions, quantum systems are in definite physical states only upon measurement.[15] It is in this sense that the trace-distance can be thought to characterise the (dis)similarity of two states. In the present case, then, $D_{\rho,\omega}$ is to be regarded as a measure of how physically similar the actual state of the subsystem at a particular time is to the equilibrium state. This is an important point which is worth re-stating: $D_{\rho,\omega}$ is a measure of the difference in the expectation values of operators on each of the two systems, which is the probability of getting different outcomes upon measurement for any given observable. This measure, I argued above, is a suitable proxy for the physical similarity between the actual state of the system at a particular time and the equilibrium state. We can now put this in a less precise but more intuitive way: the smaller the value $D_{\rho,\omega}$, the closer the subsystem is to equilibrium.

Now consider $\langle D_{\rho,\omega}\rangle_t$. Given the above, this represents how close the subsystem is to the equilibrium state *on average*. It follows that, if $\langle D_{\rho,\omega}\rangle_t$ is small, then the subsystem must spend most of its time close to equilibrium. For the moment, suppose that $\langle D_{\rho,\omega}\rangle_t$ is small in the requisite sense. What would this show?

It is clear that a small value for $\langle D_{\rho,\omega}\rangle_t$ cannot represent the subsystem monotonically tending towards, reaching and remaining in equilibrium. This much was

[13] The 'conceptually' is important here: average states, like ω, are not states on which actual measurements can be performed, of course. Note, also, that $D_{\rho,\rho\prime}$ is not a measure of the difference of *actual outcomes*.

[14] It is in this sense, vis. *any* operator, that this is a 'strong' distance measure.

[15] The central problem in the foundations of quantum mechanics is the so-called "measurement problem". Just what 'measurement' means (or *is*) is highly contentious. Here 'upon measurement' is to be understood as place holder for one's preferred solution to this problem and neutral with respect to the competing suggestions.

to be expected, of course and *this* is not problematic, for the aim was to derive the TD-like behaviour, as characterised above. The pertinent question is whether $\langle D_{\rho,\omega}\rangle_t$ being small is an adequate representation of the latter. It is and the situation is similar to the result obtained under (epislon-) ergodicity, as per the previous chatper.[16] If $D_{\rho,\omega}$ is small, then on average ρ_t must be close to the equilibrium state, ω. ρ_t can fluctuate out of equilibrium but flucuations are infrequent and have to be short lived. Thus, if we can show that $\langle D_{\rho,\omega}\rangle_t$ is small, then we will have derived TD-like behaviour via Aharonovian QSM.[17]

Before proceeding to examine whether, indeed $\langle D_{\rho,\omega}\rangle_t$ is small, a few other points are worth making. First, $\langle D_{\rho,\omega}\rangle_t$ is the time-average in the limit, $\lim_{\tau\to\infty}$. Here are, at least, two possibilities with which even a small value of $\langle D_{\rho,\omega}\rangle_t$ is, therefore, compatible:

> (Comp. 1) The subsystem taking an arbitrarily long (but finite) amount of time to reach equilibrium.
>
> (Comp. 2) The subsystem deviating arbitrarily far from equilibrium once it has reached it.

Notice it is not the case that anything here implies that it *will* take a long time to reach equilibrium nor that it *will* deviate away from it, rather if $\langle D_{\rho,\omega}\rangle_t$ is small, this does not preclude these possibilities. Moreover, these are *possibilities* with which TD-like behaviour as per Boltzmannian CSM is also compatible, so in this sense Aharonovian QSM is no worse than its classical counterpart.

Following on from the discussion in the previous chapter, 4.7.5, I propose to set aside the issue of relaxation times, i.e. I set aside (Comp. 1).

As regards (Comp. 2), the possibility is not ruled out yet, as we saw, large fluctuations must be infrequent and short-lived, if $\langle D_{\rho,\omega}\rangle_t$ is small. Aharonovian QSM improves on the Boltzmann CSM account in one respect, *ceteris paribus*. Recall that whilst Frigg & Werndl derived TD-like behaviour via epsilon-ergodicity, no further characterisation of the entropy profile of the system was given. However,

[16] Unlike the classical case, however, this results for all possible initial conditions, for the decomposition of the system, $\mathcal{H}$, into subsystem and bath was arbitrary. Put colloquially there is no Aharonovian QSM analog to the *Measure Zero* problem here. However, as discussed in the next section something analagous to the *Measure Zero* problem does appear when we consider Theorem 2.

[17] Of course, deriving this does not yet constitute a reduction of the Second Law$_S$, for one has to show that there is *warrant* for the auxiliary assumptions and bridge-law used in the derivation. I address this below.

Linden et al also provide a sharper, probabilistic characterisation of possible fluctuations to the effect that large deviations away from equilibrium are unlikely. Let us first consider the theorem itself:

Theorem 3. $\mathrm{Pr_t}\left\{\mathrm{D}_{\rho,\omega} > \sqrt{\frac{\mathrm{d_S}}{\mathrm{d}^{\mathit{eff}}(\omega_\mathrm{B})}} + \epsilon\right\} \leq \exp\left(-\mathrm{c}''\epsilon^4\mathrm{d}^{\mathit{eff}}(\omega)\right)$,

where $c'' = \frac{1}{128\pi^2}$ *and* $\epsilon << 1$.

This result holds under a different assumption about the Hamiltonian, namely that "the energy eigenvalues of [the Hamiltonian] have no rational dependencies" (Linden et al., 2009, 4) This means that the energy eigenvlaues are assumed to be *rationally independent*. Explicitly: for all the energy eigenvalues $E_1 \ldots E_n$: there are no (non-trivial, i.e. $k_i \neq 0$) rational numbers k_i such that $k_1E_1 + k_2E_2 + \cdots + k_nE_n = 0$. I return to this assumption in section 5.4.3. Bracketing that, what does this theorem tell us? It says that the probability that a fluctuation away from equilibrium of a magnitude of the square-root of the ratio of the 'size' of the subsystem to its bath is exponentially small in d^{eff}. Intuitively, it is incredibly unlikely that a small subsystem of a large system will fluctuate out of equilibrium in such a way as to be notable with respect to its bath. This is an even better characterisation of TD-like behaviour, in an obvious sense.

Crucially, the derivation of TD-like behaviour via Theorem 1 only holds conditionally; it holds *if* $\langle D_{\rho,\omega}\rangle_t$ is small. In the next section I consider Linden et al's argument for the conclusion that $\langle D_{\rho,\omega}\rangle_t$ is typically small. I will show that, as it stands, their argument for this conclusion is conceptually confused. However, one can remedy this to give the desired conclusion, as I show.

5.4.2 State-averaging and Typicality

What does it mean to say that $\langle D_{\rho,\omega}\rangle_t$ is typically small? Intuitively, it means that for the overwhelming majority of states, the value of $\langle D_{\rho,\omega}\rangle_t$ *is* small. (To say that basketball players are typically tall is just to say that the vast majority of basketball players *are* tall.) More abstractly: a particular element is typical with respect to a set of elements just when the particular element shares some property with the 'overwhelming majority' of the other elements in the set. To make this precise one needs to define the set of elements, a property, and a measure of 'size' on that set, and to set a value for what constitutes 'overwhelming majority' expressed in terms of this measure. So to say that typical basketball players are

tall then is to say that there is a set, say all the basketball players in the world, such that for the property of, say being over 190 cm, and given the measure 'number of players', ninety percent of them have that property. Typicality is also closely related to probability: again intuitively, if basketball players are typically tall then the probability that a particular basketball player is tall is high. Conversely, if the probability that basketball players are tall is high, then a particular basketball player is typically tall.[18]

Theorem 2 purportedly underpins the claim that $\langle D_{\rho,\omega}\rangle_t$ is typically small. How so? By Theorem 1, $\langle D_{\rho,\omega}\rangle_t$ is small if d^{eff} is large because $d_{\mathcal{S}}$ is small by hypothesis.[19] That Recall what Linden et al write about part (i) of Theorem 2:

> "[It] tells us that the average effective dimension is larger than half the dimension of the Hilbert subspace $[\mathcal{H}_R]$, *so when we draw states [at random] from a subspace of large dimension, the effective dimension* $d^{\text{eff}}(\omega)$ *of a typical state is large.*" (Linden et al., 2009, 5, emph. added)

So any small subsystem of a typical state will equilibrate because the effective dimension of some such state will be large, or so Linden et al. claim. However, this is in fact very confused.

The crucial ingredient here is 'state-averaging'. $\langle d^{\text{eff}}(\omega)\rangle_\Psi$ is calculated by taking the average of $d^{\text{eff}}(\omega)$ for all states $\mid \Psi(t)\rangle \in \mathcal{H}_R \subset \mathcal{H}$ by assigning a uniform distribution over $\mathcal{H}_R$ (i.e. by giving each state the same 'weight'). However, it is not the case that the property ascribed to the average of $d^{\text{eff}}(\omega)$ is also a property of a particular $d^{\text{eff}}(\omega)$ and, crucially, *not even typically!*

The contrast with Theorem 1 helps to make this clear. In Theorem 1 a property is ascribed to every state, albeit an 'average property'. Of course, recourse to an 'average property' has limitations (as discussed in section 5.4.1) but nonetheless *every* state has the property. In Theorem 2 an *average state* is ascribed a particular property, vis. being bound to the dimensionality of the Hilbert space

[18] Typicality arguments (that is: arguments that involve the notion of 'typicality') have been used in the context of statistical mechanics before. Frigg (2009) provides a precise and technical exposition of 'typicality' and details three typicality arguments for the approach to equilibrium based on classical Boltzmannian statistical mechanics. It falls beyond the scope of the present work to assess these arguments. For present purposes, however, the informal exposition, as per the above, suffices. It is important to note that Aharonovian QSM do not fall prey to the arguments against typicality approaches presented by Frigg (2009) and Frigg and Werndl (forthcoming), for Aharonovian QSM does make recourse to the dynamics of systems.

[19] As noted above, whether $d_{\mathcal{S}}$ is small in the relevant sense is an open problem.

from which the states which make up the average state are drawn. Whether any particular state has this property cannot be inferred from the theorem. And it is important to notice that typicality cannot play a role here. In the present context it has not been demonstrated that (anything like) the overwhelming number of states out of which the average is calculated have the relevant property, so a typicality argument does not even get off the ground. For example, the average height of four people may be 170cm but it does not follow that any individual has a height of (anywhere close to) 170cm. For all one can infer from the average, two of the people may be toddlers and two basketball players. Linden et al.'s claim is conceptually confused: part (i) of Theorem 2 cannot underpin the claim that $\langle D_{\rho,\omega}\rangle_t$ is typically small.

Part (ii) of Theorem 2, however, does. *Pace* Linden et al., it is not the case that part (ii) "makes the result [in part (i)] even sharper..." (Linden et al., 2009, 5); rather part (ii) is a conceptually distinct claim. Specifically, it says that for *any* state $\mid \Psi(t)\rangle \in \mathcal{H}_{\mathcal{R}}$, the probability that $d^{\text{eff}}(\omega)$ is *less than* $\frac{d_{\mathcal{R}}}{4}$ is exponentially small. And so, the probability that $d^{\text{eff}}(\omega)$ is *greater than* $\frac{d_R}{4}$ is exponentially large. So, it follows that $d^{\text{eff}}(\omega)$ is typically greater than $\frac{d_R}{4}$.

With the conceptual confusion clarified and with part(ii) of Theorem 2, we can now derive the desired result: an arbitrary small subsystem of a large quantum system *typically* tends towards and remains in, or close to, equilibrium.

1) As per section 5.4.1, if $d^{\text{eff}}(\omega)$ is large relative to $d_{\mathcal{S}}^2$, then an arbitrary small subsystem of a large quantum system tends towards and remains in, or close to, equilibrium.

2) As per section 5.4.2, for any state, $\mid \Psi(t)\rangle \in \mathcal{H}_{\mathcal{R}}$, $d^{\text{eff}}(\omega)$ is typically greater than $\frac{d_R}{4}$.

3) By hypothesis: $\frac{d_R}{4}$ is far larger than $d_{\mathcal{S}}^2$.

4) From (2) and (3): $d^{\text{eff}}(\omega)$ is *typically* large relative to $d_{\mathcal{S}}^2$.

5) From (1) and (4): an arbitrary small subsystem of a large quantum system *typically* tends towards and remains in, or close to, equilibrium.

Notice that, as with the Ergodic progam discussed in the previous chapter (in particular recall the discussion of the *Measure Zero* problem. cf. 4.7.4.1) there are two different probabilistic notions involved here. First, as we saw in section

5.4.1, if $\langle D_{\rho,\omega}\rangle_t$ is small, one gets probabilistic characterisation of the behaviour for the subsystem: it is likely that it will tend towards and then remain in or close to, equilibrium. (This is analogous to TD-like behaviour in the classical case.) Second, it is typically the case that $\langle D_{\rho,\omega}\rangle_t$ is small. Still, some systems may be atypical in the sense given above but it is very unlikely. (This is analogous to certain intial conditions (possibily) not being behaving TD-like, with probability zero, if the system is ergodic, and with probability ϵ if the system is epsilon-ergodic.)

5.4.3 *Warrant* in Aharonovian QSM

Whilst we have derived the desired result, from the point of view of Neo-Nagelian reduction, our job is not complete. In order for the derivation to constitute a reduction, one needs to show that there is *warrant* for the auxiliary assumptions and bridge-law used in the derivation.

What are the auxiliary assumptions and bridge-laws here? The important auxiliary assumption here is the dynamical one about the Hamiltonian of the system. I come back to this shortly. First let us consider the bridge-law.

Recall that the bridge-law in Aharonovian QSM is that $\omega_{\mathcal{S}}$[20] is the equilibrium state. That is, we have a bridge-law of the form $\rho_{\mathcal{S}}(E_{\mathcal{S}}) = \omega_{\mathcal{S}}$. Is this bridge-law *warranted*? First let us contrast this bridge-law with the ones encountered in CSM. The Gibbsian and Boltzmannian CSM approaches to reducing the Second Law invoked bridge-laws for entropy: in each case the relevant statistical mechanical entropy is connected with the thermodynamic entropy. As we saw, the Gibbsian bridge-law was arguably *unwarranted* because it is hard to show that there is *conceptual fit*, whereas the Boltzmannian approached faired better in this respect.

In Aharonovian QSM the bridge-law does not pertain to entropy at all but to the equilibrium state directly: it is the equilibrium state in TD which is connected to the equilibrium state in QSM. Is this bridge-law *warranted*? *By construction*, $\omega_{\mathcal{S}}$ is the (abstract) state in which the micro-properties of the system do not change over time. Assuming that the macro-properties of the system supervene on the micro-properties, it follows that $\omega_{\mathcal{S}}$ is the state in which the macro-properties of the system are unchanging too. Thus, there is at least one sense in which there is some *conceptual fit* for this bridge law: the TD equilibrium state, $\rho_{\mathcal{S}}(E_{\mathcal{S}})$ is just

[20] This was abbreviated to ω in $\langle D_{\rho,\omega}\rangle_t$.

that state where the thermodynamic properties of the system do not change over time. Of course, it may argued that this is not sufficient to show *conceptual fit*: that the properties of the system are unchanging in this state does show that this *is* an equilibrium state of the system. But this last point is best understood as putting pressure on *formal consistency* rather than *conceptual fit*, I suggest. That is, the claim that $\omega_{\mathcal{S}}$ is the equilibrium state would be supported by showing that is it *formally consistent* with the TD equilibrium state. $\omega_{\mathcal{S}}$ does not indicate the values of the macro-properties of the state (in the way that the TD equilibrium does), nor the stability of such a state, in the sense of whether it is impervious to small perturbations.[21] To show that there is *warrant* for this bridge-law would require showing that there is the requisite *formal consistency*. In particular, one would need to show how to calculate the values of thermodynamic properties like temperature and pressure in this state, for example. The question of the *warrant* for the $\rho_{\mathcal{S}}(E_{\mathcal{S}}) = \omega_{\mathcal{S}}$ bridge-law remains an open problem for Aharonovian QSM.

Let us now consider the dynamical auxiliary assumptions, viz. the assumptions about the Hamiltonians. Recall that there are two different assumptions made, one with respect to Theorems 1 and 2, and another with respect to Theorem 3. Is there *warrant* for them? In this context, the general 'test' for the *warrant* for the auxiliary assumption, namely that in making it less counterfactual one gets empirically more adequate laws, is inappropriate: Just as with the derivation of TD-like behaviour in the Boltzmannian CSM, we need the auxiliary assumptions to be true for the law one is deriving is true.

So are the auxiliary assumptions true of the relevant quantum systems? Let us consider the assumption underpinning Theorems 1 and 2 first. Linden et al emphasise that "the restriction to Hamiltonians that have no degenerate energy gaps is an extremely natural and weak restriction." (Linden et al., 2009, 2). In what sense is this an "extremely natural and weak restriction"? Linden et al offer two defences of it. First, in brackets directly after the above quote they write:

> "Indeed, adding an arbitrarily small random perturbation to any Hamiltonian will remove all degeneracies." (*ibid.*)

This argument seems ad hoc. More significantly, though, it falls prey to the "Deus Ex Machina" argument outlined in section 5.5.2: the perturbation comes from "outside" of the system and is, in this sense, non-physical.

[21] Thanks to Miklos Rédei for emphasising this latter point to me.

Second, in a discussion of the derivation of Theorem 1, they write:

> "[W]e did not assume anything special about the interaction (apart from not having degenerate energy gaps - which rules out only a set of Hamiltonians of measure zero)" (Linden et al., 2009, 4)

This too seems unpersuasive: surely all that matters is whether, for an particular system of interest, its Hamiltonian has no degenerate energy gaps? The fact that the set of such Hamiltonians are of measure zero (even setting aside the question of which measure one ought to use) in the set of all Hamiltonians is irrelevant to the evolution of the system of interest. One needs to show that the systems of interest have such Hamiltonians.

So too with the other assumption underpinning Theorem 3, namely that the energy eigenvalues of the Hamiltonian have no rational dependencies. Linden et al proffer the following:

> "Making the assumption that the eigenenergies E of H have no rational dependencies (*which is much stronger than our non-degenerate energy gaps condition*)." (Linden et al., 2009, 12, emph. added)

However, their sense of 'much stronger' is the mathematical sense: the set of Hamiltonians for which this conditions holds is a subset of the Hamiltonians of the previous kind. These are even more 'special', as it were. But again this is not the salient issue: one what needs to show is that the systems of interest have such Hamiltonians so as to derive the better-bounded version of TD-like behaviour. As with the *warrant* for the bridge-laws then, the *warrant* for these auxiliary assumptions is an outstanding issue.

To summarise: the reduction of the Second Law to quantum statistical mechanics is not complete but at least looks promising (especially in comparison to other approaches in QSM). The derivation of the theorems 1 and 2, as reconstructed at the end of 5.4.2, is an interesting result. However, to constitute a reduction, one must demonstrate that there is *warrant* for the auxiliary assumptions in this case. This is only partially successful. There is some *warrant* for the auxiliary assumptions and bridge-laws, although more work is needed to show that Aharonovian QSM affords an entirely successful reduciton of the Second Law. In particular, what needs to be shown is that there is *formal consistency* for the equilibrium bridge-law and that the Dynamical auxiliary assumptions are true of the relevant systems.

In the final section of this chapter I consider a broader, more philosophical, class of putative problems for Aharonovian QSM. To this I now turn.

5.5 Interventionism

Having set out the 'narrow' challenges facing the Aharonovian QSM apropos reduction, I now consider a 'broader' conceptual challenge. The Aharonovian QSM is clearly an interventionist approach: the system of interest (i.e. the one for which one wants to show equilibration for) is construed as a subsystem of a large quantum system, and it is in virtue of being embedded in this larger environment (or bath) that the subsystem equilibrates. However, interventionism is, in so far as it is discussed, considered to be a misguided stance in the context of the foundations of statistical mechanics. (cf. Sklar (1993); Frigg (2008); Davies (1974); Horwich (1987); Bricmont (1995); and Ridderbos and Redhead (1998).) I shall restrict our attention to this at it impinges on reduction: does the fact that Aharonovian QSM is an interventionist approach undermine the explanatory import that the reduction affords (on the supposition that it the aforementioned problems are solved)?

To be able to answer this question we must first be clear about what interventionism is. Unfortunately, there is no canonical of statement of what interventionism in the context of statistical mechanics. As a working definition of interventionism, I suggest the following:

> *The causal efficacy of the environment is a necessary condition for the exhibited behaviour of the system.*

Let me motivate this characterisation. I take it that interventionism involves more than the casual efficacy of the environment, in the sense that both the dynamics of the system of interest and the dynamics of the environment are necessary features: in general I take it, the environment would not determine the sought-after behaviour of the system of interest *irrespectively* of the dynamics of the system. On the present account, of course, the environment and the system of interest are governed by the same dynamics (namely, H) and indeed there is no 'deep fact of the matter' as to the delineation of the two (because the results hold for arbitrary decompositions of the system). Were there a result that shows that irrespectively of the target system's dynamics it would exhibit the

behaviour of interest owing to the effect of the environment, this would show the causal efficacy of the environment to be a sufficient condition. (One might then differentiate between 'weak' and 'strong' interventionism.) I further take it that any interventionist account requires the causal efficacy of the environment to be a necessary condition: if the efficacy of the environment need not feature in the account, then it just would not be an interventionist account after all.

Why is interventionism misguided? I identify three arguments in the literature that claim it is, and label them as follows: "Incredulity", "Deus ex machina" and "All for nothing". I consider each in turn, and conclude that none of them is a knock-down argument against Aharonovian QSM.

5.5.1 Incredulity

The "Incredulity" argument is sometimes heard in informal discussions of interventionism. The idea is that it is simply incredible that the causal efficacy of the environment is a necessary condition for the exhibited behaviour of the system in question. Bricmont makes this point emphatically:

> "I cannot with a straight face tell a student that (part of) our explanation for irreversible phenomena on earth depends on the existence of Sirius." (Bricmont, 1995, 199)

As it stands, this is *not even* an argument however; and one cannot, in the words of David Lewis, "refute an incredulous stare." Bricmont's sentiment seems to be that an account of the behaviour of systems ought not to depend on the environment in which the system is embedded. But why? The burden of a motivation for 'ought' lies squarely on the shoulders of the interventionist sceptic, I contend.

5.5.2 Deus Ex Machina

A more interesting argument against interventionism comes from the "Deus Ex Machina" objection:

> "A common objection against [interventionism] points out that we are always free to consider a larger system, consisting of our original system and its environment. For instance, we can consider the 'gas cum box' system... Interventionism... seems wrong because it treats

> the environment as a kind of *deus ex machina* that is somehow outside physics; but the environment is governed by the fundamental laws of physics just as the system itself is and so it cannot do the job that the interventionist has singled out for it do." (Frigg, 2008, 164-165)

Whether or not this is a persuasive argument against interventionism in general (or in CSM, the context in which Frigg is writing) is orthogonal to the current considerations. The kind of explanatory regress that is highlighted, does not take place in the Aharanov approach because the environment is *not* treated as a kind of *deus ex machina*. It is in virtue of the entire target system cum environment being governed by the same physical dynamics that the earlier results hold.

5.5.3 All for Nothing

The "All for Nothing" argument is that an interventionist stance in the foundations of statistical mechanics is all for nothing because statistical mechanics aims to account for the Second Law of thermodynamics but the Second Law only pertains to isolated systems! For example, Ainsworth (2005) writes:

> "There is a tension [with the Second law] which makes claims only about the thermodynamic entropy of closed systems and a difficulty in spelling out a modified version of the law more precisely; after all, thermodynamic entropy doesn't always increase in open systems (which is why the second law was originally formulated with respect to closed systems!)." (Ainsworth, 2005, 630)

I think that is an important observation but the following distinction ought to be made:[22] One claim is that not all non-isolated systems (in the thermodynamic sense) do exhibit equilibration. The second is whether, *ceteris paribus*, the fact that Aharonovian QSM holds for non-isolated systems undermines the putative reduction of the Second Law given that the Second Law is formulated for isolated systems.

As regards the first, not all non-isolated systems do exhibit equilibration - quite - but *it need not be in virtue of their being non-isolated that they exhibit the behaviour that they do.* This is compatible with Aharonovian QSM, as it

[22] Ainsworth did not make this remark with Aharonovian QSM to mind, of course, but the point has purchase against it.

does not make an exceptionless claim about the behaviour of systems; it claims that systems *typically* equilibrate. The challenge would then be to show why some systems are atypical in this respect, but there is no reason to think this is challenge cannot be met within the framework itself.

As regards the second: quantum mechanics tells us that entanglement is ubiquitous and that, therefore, there are no genuinely isolated systems after all. Thus, given the 'metaphysical picture' of the reducing theory, there is no such thing as a genuinely isolated system. This observation does not undermine the putative reduction but rather underscores precisely what is at stake in reduction: we want an explanation of why it is that a false to-be-reduced theory (from the point of view of the reducing one) is, nonetheless, empirically successful (at least to the extent that it is.) If we can derive the to-be-reduced theory's laws (or laws approximating them) - in this case TD-like behaviour - for just those system for which those laws hold (to the extent that they do) then we will have explained this after all. Thus the fact that, in this case, thermodynamics stipulates that the law holds for isolated systems does not undermine the putative reduction because just those systems that thermodynamics claims are isolated, are, in fact, not isolated according to the reducing theory.

5.6 Chapter Summary

In this chapter I have critically examined recent claims that the Second Law of thermodynamics can be derived from quantum statistical mechanics. I briefly showed that Albertian and H&Sian QSM are not fully-fledged accounts. I then considered Aharonovian QSM. Using the Neo-Nagelian model as a normative framework, I reconstructed the theorems which Linden et al. provided to form a partial reduction of the Second Law. Problems remain with this: the *formal consistency* of the bridge-law needs to be shown, as does the veracity of the Hamiltonian auxiliary assumptions. I hope to find solutions to these problems in future work.

Conclusions

In this thesis I have developed and defended a new model of intertheoretic reduction, Neo-Nagelian reduction. I took the derivation of the Boyle-Charles law from statistical mechanics, specifically the kinetic theory of gases, as constitutive of reduction and abstracted a model of reduction from a rational reconstruction of this case. I defended this general method as a means to avoid recourse to intuition, which, I argued, is dubious in this context.

Neo-Nagelian reduction is an exercise in explanation: the aim of reduction is an explanation of the empirical success of the reduced theory by the reducing theory.

Reduction consists in deriving the laws of the to-be-reduced theory from the reducing theory, and auxiliary assumptions and bridge-laws. Reduction *qua* explanation is underpinned by the DN-model of explanation, which I rehabilitated in the first chapter, and deriving the laws of the to-be-reduced theory is an explanation of the theory's explanatory success precisely because it is its laws that encode its empirical content. However, the mere derivation of the laws of the to-be-reduced theory is not an explanation of the empirical success of the theory, as I have been at pain to stress. Indeed, I think that the most significant flaw with Nagel's model, even in its best version: the Schaffner-modified Nagelian model, is the failure to recognise this point. Philosophers of science have too long fixated on merely *deriving* the laws of the theories to be reduced. In contrast, I argued that in order to afford explanation, the auxiliary assumptions and bridge-laws from which they are derived need to be explanation supporting; put coarsely, not any old derivation will do!

I conceptualised the sense in which the auxiliary assumptions and bridge-laws are (need to be) explanation supporting with the concept of *warrant*. The most important kinds of auxiliary assumption are the dynamical and idealising ones.[23]

[23]There are various kind of auxiliary assumptions, no doubt not exhausted by the kinds I

What counts as *warrant* for such auxiliary assumptions? A crucial point is this: the *warrant* for such auxiliary assumptions is not, in general, determined by their counterfactualness. Failure to appreciate this point has had a detrimental affect on discussions about reduction; how successful a reduction is has often been confused with how empirically successful the to-be-reduced theory is. These are two different things. The measure of the empirical adequacy of a theory can be measured by the counterfactualness of the auxiliary assumptions used to derive its laws. The more counterfactual they are the less empirically adequate the theory is bound to be. However, the success of a reduction, i.e. *how good an explanation of the empirical adequacy of the to-be-reduced theory one has*, cannot be measured by the same thing. Rather, I have argued, such auxiliary assumptions are *warranted* in so far as one finds an affirmative answer to the following question: does making the auxiliary assumptions less counterfactual yield empirically more adequate laws? If so, the auxiliary assumptions are, so to speak, hooked-up to the world in the right way. However, there is a limiting case, as the examples considered in chapters 4 and 5, namely the reduction of the Second Law of thermodynamics to Boltzmannian classical mechanics and Aharanovian quantum statistical mechanics, showed. Obviously, where the law to be derived is true (not just approximately empirically adequate) then the dynamical auxiliary assumptions involved in such a derivation need to be true too.

As regards bridge-laws, I have argued that they are a particular kind of theoretical stipulation, namely *coherence constraints*. The *warrant* for bridge-laws comes from showing both *formal consistency* and *conceptual fit*. *Formal consistency* amounts to the requirement that the property in the reducing theory is formally equivalent to the property in the to-be-reduced theory with which the bridge-law connects it to. For example, we saw that there is perfect *formal consistency* between the thermodynamic entropy and the Gibbs entropy, for the latter 'behaves' exactly like the former. The Boltzmann entropy, at least in the case of gases, also 'behaves' in the right way. In contrast it is not clear whether there is *formal consistency* between the thermodynamic equilibrium state and the Aharanovian QSM equilibrium state. However, *formal consistency* is not enough. To be *warranted* bridge-laws also need *conceptual fit*. This is a rather textured and context specific notion; necessary and sufficient conditions are not forthcoming. I

have considered. I hope that my characterisation of the concept of *warrant* for the ones I have considered is suggestive enough for the concept to be extended as required.

think, however, that the notion has intuitive appeal as the examples encountered show. First recall the temperature case: for a specific isolated system, temperature is directly proportional to the internal energy of the system and this fits with the the statistical mechanical description ('metaphysical picture') of the system in terms of mean kinetic energy. Now reconsider the thermodynamic and Gibbs entropies. Here there is no conceptual fit for the latter is not even a property of the relevant system. There is excellent *conceptual fit* between the thermodynamic and Aharanovian QSM equilibrium states on the other hand. The Aharanovian QSM equilibrium state is (indeed by construction) the state in which the properties of the system are unchanging, which is exactly how the equilibrium state in thermodynamics is conceptualised.

Much of the literature on reduction has been concerned with ontological simplification. What I called the *Simplistic Ontological Simplification* thesis is as often contested as it is advocated but it appears in almost all discussions about reduction. To my mind these debates are misguided. Whilst ontological simplification is not a necessary condition for Neo-Nagelian reduction, a successful reduction does afford ontological simplification and, importantly, does so in a far more metaphysically tempered way. Adopting Quine's meta-ontological position, I argued that a successful reduction offers good reason to exclude the reduced theory from our best conceptual scheme and, if so, we are not committed to its ontology. Quine's meta-ontological position, whilst appealing I contend, is not uncontroversial. More needs to be said in its defence for the Neo-Nagelian account of ontological simplification to be shown to be sound.

With that last point in mind, let me then finish by sketching future areas of research. I hope to further research the prospects of Aharanovian QSM. In particular, using the Neo-Nagelian model as a normative framework, it is interesting to consider the *formal consistency* between the thermodynamic and Aharanovian QSM equilibrium state in more detail. Indeed, Aharanovian QSM in general is ripe for philosophical inquiry, especially with an eye to the question of reduction.

I would also like to apply the Neo-Nagelian model to other pairs of theories. In the first instance it would be interesting to apply it to 'classic' cases such as, in order of ambitiousness!, Keplerian and Newtonian celestial mechanics, Newtonian mechanics and relativity theory, and classical and quantum mechanics. Second, I think it would be fruitful to consider cases where the to-be-reduced and/or reducing theories are not fully formalised. Can the Neo-Nagelian model be gen-

eralised to deal with such cases as the relation between biology and chemistry, and psychology and neuroscience?

Finally, what of emergence? 'Emergence' is eerily absent throughout this thesis. It is often considered the counterpart of reduction: if one theory does not reduce to another, then it is an emergent theory. It was tempting to offer a conception of emergence in terms of the Neo-Nagelian account of reduction. However, in the end I could not do justice to it here; I hope to do so in the future.

Bibliography

Ager, T. A., Aronson, J. L., and Weingard, R. (1974). Are Bridge Laws Really Necessary? *Noûs*, 8(2): 119–134.

Ainsworth, P. M. (2005). The Spin-Echo Experiment and Statistical Mechanics. *Foundations of Physics Letters*, 18(7): 621–635.

Albert, D. Z. (1994). The Foundations of Quantum Mechanics and the Approach to Thermodynamic Equilibrium. *British Journal for the Philosophy of Science*, 45(2): 669–677.

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

Azzouni, J. (1998). On "On What There Is". *Pacific Philosophical Quarterly*, 79(1): 1–18.

Balzer, W., Moulines, C., and Sneed, J. (1987). *An architectonic for science: the structuralist program*. Synthese library. D. Reidel Pub. Co.

Batterman, R. W. (1995). Theories Between Theories: Asymptotic Limiting Intertheoretic Relations. *Synthese*, 103(2).

Batterman, R. W. (2000). Multiple Realizability and Universality. *British Journal for the Philosophy of Science*, 51(1): 115–145.

Bechtel, W. P. and Mundale, J. (1999). Multiple Realizability Revisited: Linking Cognitive and Neural States. *Philosophy of Science*, 66(2): 175–207.

Bickle, J. (1996). New Wave Psychophysical Reductionism and the Methodological Caveats. *Philosophy and Phenomenological Research*, 56(1): 57–78.

Bickle, J. (1998). *Psychoneural Reductionism: The New Wave*. MIT Press.

Bishop, R. C. and Atmanspacher, H. (2006). Contextual Emergence in the Description of Properties. *Foundations of Physics*, 36(12): 1753–1777.

Block, N. and Fodor, J. A. (1972). What Psychological States Are Not. *Philosophical Review*, 81(April): 159–81.

van Bouwel, J. and Weber, E. (2008). A Pragmatist Defense of Non-Relativistic Explanatory Pluralism in History and Social Science. *History and Theory*, 47(2): 168–182.

Brandt, R. and Kim, J. (1967). The Logic of the Identity Theory. *Journal of Philosophy*, 66(September): 515–537.

Bricmont, J. (1995). Science of Chaos or Chaos in Science? *Physicalia Magazine*, 17: 159–208.

Brown, H. R. and Uffink, J. (2001). The Origins of Time-Asymmetry in Thermodynamics: The Minus First Law. *Studies in History and Philosophy of Science Part B*, 32(4): 525–538.

Brush, S. (1986). *The kind of motion we call heat: a history of the kinetic theory of gases in the 19th century*. Number v. 2 in North-Holland personal library. North-Holland.

Callender, C. (1999). Reducing Thermodynamics to Statistical Mechanics: The Case of Entropy. *Journal of Philosophy*, 96(7): 348–373.

Callender, C. (2001). Taking Thermodynamics Too Seriously. *Studies in History and Philosophy of Science Part B*, 32(4): 539–553.

Carnap, R. (1950). Empiricism, Semantics, and Ontology. *Revue Internationale De Philosophie*, 4(2): 20–40.

Carnap, R. (1967). *The Logical Structure of the World [and] Pseudoproblems in Philosophy*. London, Routledge K. Paul.

Causey, R. L. (1972). Attribute Identities in Microreductions. *Journal of Philosophy*, 64(August): 407–22.

Chakravartty, A. (2001). The Semantic or Model-Theoretic View of Theories and Scientific Realism. *Synthese*, 127(3).

Chang, H. (2007). *Inventing Temperature: Measurement and Scientific Progress*. Oxford Studies in the Philosophy of Science. Oxford University Press. ISBN 9780195337389.

Chen, Z. (2004). Quantum Theory for the Binomial Model in Finance Theory. *Journal of Systems Science and Complexity*, 17: 567–573.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. The Mit Press.

Churchland, P. M. (1979). *Scientific Realism and the Plasticity of Mind*. Cambridge University Press.

Churchland, P. M. (1985). Reduction, Qualia and the Direct Introspection of Brain States. *Journal of Philosophy*, 82(January): 8–28.

Churchland, P. S. (1986). *Neurophilosophy: Toward A Unified Science of the Mind-Brain*. MIT Press.

Clapp, L. J. (2001). Disjunctive Properties: Multiple Realizations. *Journal of Philosophy*, 98(3): 111–136.

Davies, P. (1974). *The Physics of Time Asymmetry*. University of California Press.

Dugdale, J. (1996). *Entropy and its Physical Meaning*. Taylor and Francis.

Earman, J. and Rédei, M. (1996). Why Ergodic Theory Does Not Explain the Success of Equilibrium Statistical Mechanics. *British Journal for the Philosophy of Science*, 47(1): 63–78.

Eck, D. V., Jong, H. L. D., and Schouten, M. K. D. (2006). Evaluating New Wave Reductionism: The Case of Vision. *British Journal for the Philosophy of Science*, 57(1): 167–196.

Enc, B. (1983). In Defense of the Identity Theory. *Journal of Philosophy*, 80(May): 279–98.

Endicott, R. P. (1998). Collapse of the New Wave. *Journal of Philosophy*, 95(2): 53–72.

Endicott, R. P. (2001). Post-Structuralist Angst - Critical Notice: John Bickle, _Psychoneural Reduction: The New Wave_. *Philosophy of Science*, 68(3): 377–393.

Endicott, R. P. (2005). Multiple Realizability. In *Encyclopedia of Philosophy, 2nd edition*. Thomson Gale, Macmillan Reference.

Esfeld, M. and Sachse, C. (2007). Theory Reduction by Means of Functional Sub-Types. *International Studies in the Philosophy of Science*, 21(1): 1 – 17.

Feyerabend, P. K. (1965). On the "Meaning" of Scientific Terms. *Journal of Philosophy*, 62(10): 266–274.

Fodor, J. A. (1974). Special Sciences. Or: The Disunity of Science As A Working Hypothesis. *Synthese*, 28: 97–115.

Førland, T. E. (2004). The Ideal Explanatory Text in History: A Plea for Ecumenism. *History and Theory*, 43(3): 321–340.

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

Friedman, M. (1974). Explanation and Scientific Understanding. *Journal of Philosophy*, 71(1): 5–19.

Frigg, R. (2008). A Field Guide to Recent Work on the Foundations of Statistical Mechanics. In D. Rickles, editor, *The Ashgate Companion to Contemporary Philosophy of Physics*, Ashgate companion. Ashgate Pub. Ltd.

Frigg, R. and Werndl, C. (forthcoming). Explaining Thermodynamic-Like Behaviour in Terms of Epsilon-Ergodicity. *Philosophy of Science*.

Ghirardi, G., Rimini, A., and Weber, T. (1986). Unified Dynamics for Microscopic and Macroscopic Systems. *Phys. Rev. D*, 34: 470.

Giere, R. N. (1994). The Cognitive Structure of Scientific Theories. *Philosophy of Science*, 61(2): 276–296.

Gillett, C. (2003). The Metaphysics of Realization, Multiple Realizability, and the Special Sciences. *Journal of Philosophy*, 100(11): 591–603.

Glymour, C. (1970). On Some Patterns of Reduction. *Philosophy of Science*, 37(3): 340–353.

Goldstein, S. and Lebowitz, J. L. (2004). On the (Boltzmann) Entropy of Nonequilibrium Systems. *Physica*, D: 53 – 66.

Hecht, C. (1998). *Statistical Thermodynamics and Kinetic Theory*. Dover books on physics. Dover Publications.

Hemmo, M. and Shenker, O. (2001). Can We Explain Thermodynamics By Quantum Decoherence? *Studies in History and Philosophy of Science Part B*, 32(4): 555–568.

Hemmo, M. and Shenker, O. (2003). Quantum Decoherence and the Approach to Equilibrium. *Philosophy of Science*, 70(2): 330–358.

Hitchcock, C. (2003). Unity and Plurality in the Concept of Causation. In F. Stadler, editor, *The Vienna Circle and Logical Empiricism*, volume 10 of *Vienna Circle Institute Yearbook*, pp. 217–224. Springer Netherlands.

Hooker, C. A. (1981). Towards a General Theory of Reduction. Part II: Identity in Reduction. *Dialogue*, 20(02): 201–236.

Horgan, T. E. (1993). Nonreductive Materialism and the Explanatory Autonomy of Psychology. In S. Wagner and R. Wagner, editors, *Naturalism: A Critical Appraisal*. University of Notre Dame Press.

Horwich, P. (1987). *Asymmetries in Time: Problems in the Philosophy of Science*. MIT Press.

Huang, K. (1987). *Statistical mechanics*. Wiley.

Jackson, F. and Pettit, P. (1992). In Defense of Explanatory Ecumenism. *Economics and Philosophy*, 8(01): 1–.

Joos, E. and Zeh, D. (1985). The Emergence of Classical Properties Through Interaction with the Environment. *Zeitschrift fur Physik B*, 59: 223–243.

Kemeny, J. G. and Oppenheim, P. (1956). On Reduction. *Philosophical Studies*, 7(1-2).

Kim, J. (1992). Multiple Realization and the Metaphysics of Reduction. *Philosophy and Phenomenological Research*, 52(1): 1–26.

Kim, J. (1995). Mental Causation: What? Me Worry? *Philosophical Issues*, 6: 123–151.

Kim, J. (2000). *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. MIT Press.

Kitcher, P. and Salmon, W. (1989). *Scientific Explanation*. Univ of Minnesota Pr.

Lebowitz, J. L. (1993). Macroscopic Laws, Microscopic dynamics, Time's Arrow and Boltzmann's Entropy. *Physica A*, 194: 1 – 27.

Lebowitz, J. L. (1999). Statistical Mechanics: A selective Review of Two Central Issues. *Reviews of Modern Physics*, 71: 346 – 357.

Lewis, D. (1969). Review of "Art, Mind, and Religion". *Journal of Philosophy*, 66.

Linden, N., Popescu, S., Short, A., and Winter, A. (2009). Quantum Mechanical Evolution Towards Equilibrium. *Phys. Rev. E.*, 79: 061103.

Lyre, H. (2009). The "Multirealization" of Multiple Realizability. In A. Hieke and H. Leitgeb, editors, *Reduction - Abstraction - Analysis. Proceedings of the 31th International Ludwig Wittgenstein-Symposium in Kirchberg*. Ontos Verlag.

Marras, A. (2002). Kim on Reduction. *Erkenntnis*, 57(2): 231–57.

Mayr, D. (1976). Investigations of the Concept of Reduction I. *Erkenntnis*, 10(3): 275–294.

Mirowski, P. (1991). *More heat than light: economics as social physics, physics as nature's economics*. Historical perspectives on modern economics. Cambridge University Press.

Müller, I. (2007). *A History of Thermodynamics: the Doctrine of Energy and Entropy*. Springer.

Nagel, E. (1949). The Meaning of Reduction in the Natural Sciences. In R. Stauffer, editor, *Science and Civilization*, pp. 98–123. Madison, Wis.

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

Nagel, E. (1979). *Teleology Revisited and Other Essays in the Philosophy and History of Science*. The John Dewey essays in philosophy. Columbia University Press.

Needham, P. (2010). Nagel's Analysis of Reduction: Comments in Defense as Well as Critique. *Studies in History and Philosophy of Science Part B*, 41(2): 163–170.

Nickles, T. (1973). Two Concepts of Intertheoretic Reduction. *Journal of Philosophy*, 70(April): 181–201.

Pippard, A. (1957). *Elements of classical thermodynamics for advanced students of physics*. Cambridge University Press.

Polger, T. W. (2008). Two Confusions Concerning Multiple Realization. *Philosophy of Science*, 75(5): 537–547.

Popescu, S., Short, A., and Winter, A. (2006). Entanglement and the Foundation of Statistical Mechanics. *Nature Physics*, 21(11): 754–758.

Putnam, H. (1967). Psychological Predicates. In W. H. Capitan and D. D. Merrill, editors, *Art, Mind and Religion*. Pittsburgh University Press, Pittsburgh.

Quine, W. V. (1948). On What There Is. *Review of Metaphysics*, 2: 21–38.

Rantala, V. (1991). Review. *Synthese*, 86(2).

Reiss, H. (1996). *Methods of thermodynamics*. Dover books on physics. Dover Publications.

Richardson, R. C. (2008). Autonomy and Multiple Realization. *Philosophy of Science*, 75(5): 526–536.

Ridderbos, T. M. and Redhead, M. L. G. (1998). The Spin-Echo Experiments and the Second Law of Thermodynamics. *Foundations of Physics*, 28: 1237–1270.

Salmon, W. C. (1985). Conflicting Conceptions of Scientific Explanation. *Journal of Philosophy*, 82(11): 651–654.

Schaffner, K. F. (1967). Approaches to Reduction. *Philosophy of Science*, 34(2): 137–147.

Schaffner, K. F. (1993). Theory Structure, Reduction, and Disciplinary Integration in Biology. *Biology and Philosophy*, 8(3).

Schouten, M. K. D. and de Jong, H. L. (1998). Defusing Eliminative Materialism: Reference and Revision. *Philosophical Psychology*, 11(4): 489–509.

Shapiro, L. A. (2000). Multiple Realizations. *Journal of Philosophy*, 97(12): 635–654.

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

Sober, E. (1999). The Multiple Realizability Argument Against Reductionism. *Philosophy of Science*, 66(4): 542–564.

Suppe, F. (1974). *The Structure of Scientific Theories*. Urbana, University of Illinois Press.

Suppes, P. (1960). A Comparison of the Meaning and Uses of Models in Mathematics and the Empirical Sciences. *Synthese*, 12(2-3): 287–301.

Tolman, R. (1938). *The principles of statistical mechanics*. Dover books on physics and chemistry. Dover Publications.

Uffink, J. (2001). Bluff Your Way in the Second Law of Thermodynamics. *Studies in History and Philosophy of Science Part B*, 32(3): 305–394.

Uffink, J. (2007). Compendium of the Foundations of Classical Statistical Physics. In J. Butterfield and J. Earman, editors, *Philosophy of Physics*, pp. 923–1047. Amsterdam: North Holland.

Woodward, J. (2010). Scientific Explanation. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. The Metaphysics Research Lab, spring 2010 edition.

Wright, C. D. (2000). Eliminativist Undercurrents in the New Wave Model of Psychoneural Reduction. *Journal of Mind and Behavior*, 21(4): 413–436.

Zemansky, M. and Dittman, R. (1981). *Heat and thermodynamics: an intermediate textbook*. McGraw-Hill.

Zurek, W., Habib, S., and Paz, J. (1993). Coherent States via Decoherence. *Physical Review Letters*, 70: 1187–1190.