INELIMINABLE IDEALIZATIONS, PHASE TRANSITIONS, AND IRREVERSIBILITY

DISSERTATION

Presented in Partial Fulfillment of the Requirements for

the Degree Doctor of Philosophy in the

Graduate School of The Ohio State University

By

Nicholaos John Jones, B. A., M. A.

* * * * *

The Ohio State University

2006

Dissertation Committee:

Professor Robert Batterman, Co-Adviser

Professor Neil Tennant, Co-Adviser

Professor Stewart Shapiro

Professor Robert Perry

Approved by

____

Co-Adviser

____

Co-Adviser
Graduate Program in Philosophy

ABSTRACT

The dissertation examines two putative explanations from statistical mechanics with the aim of understanding the nature and role of idealizations in those accounts, namely, the Yang-Lee account of phase transitions and the Boltzmannian account of irreversible behavior. Like most explanations in physics, these accounts involve idealizations to some extent. Many idealized explanations hold out the hope that the idealizations can be removed or eliminated with further work. However, the idealizations that occur in the accounts of phase transitions and irreversibility are ineliminable. The only way (in principle) to obtain a description – let alone an explanation – of these phenomena is to invoke various idealizing assumptions.

Ineliminably idealized explanations are not well-understood from a philosophical point of view. Indeed, most philosophers of science would probably hold that no idealizations are ineliminable. The dissertation argues that this view is mistaken, showing where and why extant accounts of idealization miss this fact by distinguishing the widely-accepted understanding of idealizations as falsehoods from a novel understanding of idealizations as abstractions. As abstractions, idealizations are devices for ignoring certain details about the real world. The dissertation argues that ineliminable idealizations cannot be falsehoods, and that they should be understood as abstractions.

The dissertation also examines the confirmation of idealized hypotheses and their role as guides to what the world is like. At least some idealized hypotheses have some degree of confirmation; and less idealized hypotheses tend to be better confirmed than their more idealized counterparts. If idealizations are falsehoods, Bayesian confirmation theory seems unable to obtain these results, because it lacks a way of defining the prior probabilities of idealized hypotheses. If idealizations are abstractions, however, idealized hypotheses about a system are incomplete claims that omit certain details about the system. Since prior probabilities are assigned to such hypotheses in the same way they are assigned to incomplete descriptions, understanding idealizations as abstractions allows Bayesianism to secure the above-mentioned results. This understanding of idealizations also allows idealized hypotheses to be guides to what the world is like, because the incompleteness of such hypotheses is compatible with the cogency of inference to the best explanation.

To my mother

ACKNOWLEDGMENTS

This dissertation is the result of a prolonged period of massive confusion and panic. I am grateful to Bob Batterman for his patience and guidance through those confusions which I have managed to overcome. I could not have written this dissertation, nor written it as well as I have, without his extensive feedback and help.

I am indebted to Neil Tennant and Stewart Shapiro for their careful scrutiny of my work and their many suggestions on how to improve it. Neil deserves a special further mention for agreeing to be my official advisor after Bob's departure to Canada. I also thank Robert Perry for his generous assistance in helping me to understand quantum mechanics, for happily agreeing to be part of my defense committee, and for his enthusiasm in discussing philosophy and physics.

I thank George Schumm, Bill Melanson, Adam Podlaskowski, and Makoto Suzuki for helpful feedback and conversations about various portions of Chapter One. I thank Bill Roche for reading and commenting upon versions of Chapters Six and Seven, as well as for stimulating suggestions on very early portions of Chapter Four. I thank Charles Hermes for helpful comments on ideas contained in Chapter Five. I thank Joshua Smith for indulging my forays into epistemology and for all his help with LaTeX. I thank Rick Groshong, Ryan Jordan, Salvatore Florio, Cathal Woods, Henry Pratt, Mike Jaworski, Scott Bridges, Ian Smith, Andy Arlig, Patrick Matthews, and those already mentioned for making being in graduate school easier to bear in ways too numerous to mention. Bill Melanson's extraordinary 1998 "paper" also played a role that no words can describe.

I also thank: David Merli and Ryan Nichols, for their friendliness during my initial years as a graduate student and the advice they readily shared about how to make it through the program; Robert Kraut and Allan Silverman, for their support through graduate school - Robert for sharing his passion for teaching with me, Allan for his once-constant "how's the philosophy of biology doing?" queries; William Taschek, for his intermittent (but always timely) encouragement to keep at work on the dissertation, and for his asking at various points how my thoughts were coming along; Lisa Shabel, for her efforts in preparing me to go on the job market; Debbie Blickensderfer, Michelle Brown, and Sue O'Keefe, for their assistance and friendliness throughout graduate school; and the Department of Philosophy, for its generous financial support and collegial atmosphere.

I am indebted to Robert Stith, G. Steven Neeley, and Timothy Menta, teachers from my days as an undergraduate, for their guidance and encouragement during my graduate career, and for getting me excited about philosophy in the first place. I am grateful to Timothy and Melissa, my brother and sister, for being understanding at my absence from various events throughout my college days and just plain caring, and most of all to Samantha Makar, for her confidence and support through the latter stages of dissertation writing, for her affection, and for keeping my life fiercely interesting. Finally, I thank my mother, Marcella, for her constant encouragement throughout all stages of my academic career.

VITA

2001-present	Graduate Teaching Associate, The Ohio State University
2000-2004	M.A. Philosophy The Ohio State University
1995-2000	B. A. Philosophy Saint Francis College (PA)
August 4, 1977	Born, Baltimore, Maryland

PUBLICATIONS

Research Publications

Jones, Nicholaos John (2004). The Logic of *Soku* in the Kyoto School, *Philosophy East and West*, 54.3: 302 - 321.

FIELDS OF STUDY

Major Field: Philosophy

TABLE OF CONTENTS

Page

LIST OF FIGURES

CHAPTER 1

INTRODUCTORY REMARKS ON EXPLANATION, IDEALIZATION, AND INELIMINABILITY

For the most part, scientific explanations are answers to why-questions. A scientific explanation of some phenomenon P answers the question 'Why is it the case that P obtains?' One might ask why straight sticks appear to be bent when immersed in water, why some metals become magnetic, why the moon eclipses the sun, etc. Science aims to answers these questions. This explanatory task goes beyond the descriptive task of science, because explanations do not merely describe what is the case; they also show why it is the case.

A scientific explanation can be thought of as a set of sentences, one of which – the *explanandum* – stands in a certain relation to the others – the *explanans*. The explanandum is a sentence to the effect that the phenomenon to be explained obtains; there is always a presupposition that the explanandum is true, because there is no explanation of why a false sentence is true nor of why a phenomenon that does not obtain obtains. The explanans is the set of sentences that purport to explain the explanandum. Whether an explanans explains its explanandum depends upon whether there is an appropriate relationship between them; this relationship distinguishes explanations from non-explanations, preventing all but a privileged few answers to a why-question from being explanatory.

Thinking of scientific explanation in this way overlooks some issues about scientific explanation that happen to be tangential to the focus of this dissertation. For instance, there has been some debate about whether sentences, or only the facts that sentences describe, do explanatory work. No substantial thesis in this dissertation hinges upon which of these views is correct. For this reason, I take the liberty of sometimes treating a member of the explanans as a fact rather than a sentence describing the fact. I also assume the irrelevance of any difference between explaining why a sentence is true and why the phenomenon described by the sentence obtains.

Technical niceties aside, there are two putative scientific explanations on which this dissertation focuses. The first is the statistical mechanical account of the occurrence of phase transitions. This account applies to phenomena such as the melting of ice and the spontaneous magnetization of iron. The second putative explanation is a statistical mechanical account of why non-equilibrium systems, when left to themselves, irreversibly approach a state of equilibrium in a finite amount of time. This account applies to phenomena such as the spread of a puff of cigarette smoke throughout a room. A presumption of this dissertation, defended in Chapter Five, is that these accounts are genuinely explanatory.

Like many scientific explanations, the statistical mechanical accounts of phase transitions and irreversibility are idealized. An *idealized explanation* is an explanation in which at least one member of the explanans is an idealization. There are two characteristic functions of idealizations (see [49], pp. 177-178). First, every idealization replaces a description of a system with a description of an idealized version of that system. (The idealization itself is neither the replaced description nor the replacement description.) Second, every idealization simplifies: the replacement description is, in some sense, simpler than the original description.[1] For example, working with the replacement description might simplify the mathematical analysis of the system. The simplification function distinguishes idealizations from other kinds of replacement, such as the replacement of truth with fiction that occurs in some cases of deception. Note that this characterization of idealizations appeals only to the operational role of certain syntax; it appeals to neither the syntactic form of idealizations nor a semantic interpretation of such syntax.

Of all the idealizations that occur in the explanans for the accounts of phase transitions and irreversibility, an idealization common to both accounts – and the idealization most salient to this dissertation – is the limit in which a system's number N of particles "tends to infinity": $N \to \infty$. If idealizations are statements (declarative sentences), this syntax should be interpreted as being the limit in which a system's number of particles becomes infinitely large, and the application of this limit to an equation that describes a real system with only finitely many particles yields an equation that describes a (non-real) system with infinitely many particles.

Although the accounts of phase transitions and irreversibility are similar to many other scientific explanations in being idealized explanations, one thesis of this dissertation is that the accounts are dissimilar to other such explanations in being ineliminably idealized. (An idealized explanation of some phenomenon is *ineliminably*

[1]The sense in which idealizations simplify is controversial. The discussion to follow relies upon intuitions about when one description is simpler than another. For one suggested analysis, see [111].

idealized if, in principle, the only way to explain the phenomenon is to appeal to an idealization in the explanans.)

> *The statistical mechanical accounts of phase transitions and irreversible behavior are ineliminably idealized.*

Most idealized explanations are not ineliminably idealized: with most such explanations, it is possible to alter an idealized explanans for some phenomenon into an explanans that is less idealized, in such a way that the altered explanans also explains the original phenomenon. For instance, it is possible to explain why a certain launched projectile follows a parabolic path under the idealization that the projectile is a perfect sphere; and it is also possible to explain this fact without idealizing the shape of the projectile (although the latter explanation is more complicated). The interesting feature of the statistical mechanical accounts of phase transitions and irreversible behavior is that the idealizing limit in which a system's particle number $N \to \infty$ is ineliminable to those accounts.

(The arguments in favor of the ineliminability of the $N \to \infty$ limit in the accounts of phase transitions and irreversibility are based upon technical impossibility results from the current state of science. I do not mean to claim that scientific advances will never show the limit to be eliminable in accounting for these phenomena. Rather, I aim to show that, so far as we know at present, the limit is ineliminable to the accounts of phase transitions and irreversibility. Moreover, even if future progress provides explanations of these phenomena that do not require an appeal to the $N \to \infty$ limit, that progress would not undermine the thesis that the current accounts of phase transitions and irreversibility do require an appeal to this limit.)

A second thesis of the dissertation is that many extant philosophical accounts of scientific explanation fail to accommodate ineliminably idealized explanations.

Many extant philosophical accounts of scientific explanation fail to accommodate ineliminably idealized explanations.

The overarching goal of this dissertation is to provide a philosophical account of how these explanations can be explanatory despite being ineliminably idealized. But first, it will be helpful to discern the conditions a philosophical account of explanation must satisfy in order to accommodate idealized explanations in general. For if a philosophical account of scientific explanation disallows the existence of idealized explanations, it also disallows the existence of ineliminably idealized explanations.

1.1 Explanation and Distorting Idealizations

Philosophical accounts of scientific explanation attempt to provide conditions under which an answer to a why-question is explanatory. That is, they attempt to provide conditions that an explanans must satisfy if it is to explain some explanandum. Historically, these accounts take one of three forms. Nomothetic accounts, best represented by Carl Hempel's deductive-nomological model, hold that to explain a phenomenon is to subsume it under appropriate laws. Causal accounts, best represented by Wesley Salmon, hold that to explain a phenomenon is to show how the causal-nomic structure of the world produces the phenomenon. Unification accounts, best represented by Philip Kitcher, hold that to explain a phenomenon is to unify it with other phenomena. The representative versions of each of these accounts, along with a fairly prevalent semantic interpretation of idealizations, entail that there are no idealized explanations.

1.1.1 Representative Accounts of Explanation

Carl Hempel's deductive-nomological (DN) model is a representative nomothetic account of explanation. According to Hempel, "the principal requirement for scientific explanation is . . . the inferential subsumption of the explanandum under comprehensive general principles" ([43], p. 445). Hempel provides four adequacy conditions for an explanation. First, the explanans must have empirical content; its elements must be confirmable. Second, the explanans must deductively entail the explanandum. Third, at least one member of the explanans must be a statement of a law of nature, and this law-statement must be essential to the valid derivation of the explanandum. (The difficulty of adequately characterizing the notion of a law of nature, as well as the debate over whether there are such laws, are acknowledged but passed over as irrelevant to the present discussion: see [95] and [122], respectively.) Finally, every member of the explanans must be true.

Hempel takes this last condition, the factual correctness condition, to be "obvious" ([45], p. 322). He prefers an explanans that is true to an explanans that is highly confirmed, in order to prevent the possibility of an argument that satisfies his first three conditions being an explanation at some point in time but, given scientific advances that highly disconfirm some elements of the explanans, the same argument not being an explanation at some later time. Presumably his conception of explanations as sound arguments explains why he prefers true explanans to false ones.

Wesley Salmon offers a representative version of a causal account of explanation that differs from Hempel's DN model in not requiring explanations to be arguments. According to Salmon, explanations are assemblages of facts that fit the phenomenon to be explained into its causal nexus – that is, etiological facts about the causal interactions and processes that produce the phenomenon as well as constitutive facts about the composition of the system in which the explanandum occurs. This assemblage might be an argument; and it might not be. There need not be any entailment relation between the sentences that describe the way in which the explanandum is embedded in its causal nexus and the explanandum itself; the only relations required are causal relations between the facts described by the explanans and explanandum. (The details of Salmon's theory of causation are not important for the present discussion: see [103], pp. 253-257.)

Salmon's account shares with Hempel's model a requirement of factual correctness. No causal process or causal interaction is part of a causal explanation of an event unless it is part of the causal nexus for that event. Since non-existent causal processes and non-obtaining causal interactions are not part of the causal nexus for any event to be explained, they are not part of any causal explanation. David Lewis's account of causal explanation, the major competitor to Salmon's, also contains something like a requirement of factual correctness. Lewis is unwilling to decide whether accounts that violate this requirement are non-explanatory or just bad explanations: "it is unclear – and we needn't make it clear – what to say about an unsatisfactory chunk of explanatory information, say one that is incorrect or too small to suit us. We may call it a bad explanation, or no explanation at all" ([70], p. 218).

Like Hempel's and Salmon's accounts, Kitcher's account of explanation as unification (see [56], [57]) denies that falsehoods are explanatory. According to Kitcher, explanations are valid deductive arguments, and whether an argument is an explanation depends upon whether it instantiates an explanatory argument pattern.

The explanatory argument patterns are the ones that best unify the set of sentences that belong to the belief corpus of scientific practice in the limit of its rational development ([57], p. 498). Kitcher's account for when an argument pattern best unifies a belief corpus is complex. Although the details are interesting, the important point for this discussion is that his account prohibits falsehoods from being explanatory. An argument is a candidate for being among the best unifiers of a belief corpus only if all of its premises are members of that corpus ([57], p. 434). Since Kitcher identifies truths as those sentences that belong to the belief corpus of science in the limit of its rational development, no falsehood is a premise among the arguments that best unify that belief corpus.

1.1.2 Idealizations as Distortions

These accounts of scientific explanation are incompatible with the existence of idealized explanations, if the correct semantic interpretation of idealizations is that they are distortions. A *distortion* attributes a feature to a system that the system does not have. Many kinds of sentences, such as lies and ordinary mistakes, are distortions. For example, if Jack tells Jill that he cannot visit her because he has work to do over the weekend, but the reason he cannot visit her is that he has more important plans, then Jack's statement is a distortion, because it attributes to Jack a reason that he does not have. Ordinary mistakes also qualify as distortions. For instance, Newtonian mechanics allows information transfer to occur at speeds well beyond the speed of light. This is a distortion with respect to how fast information can be transferred, because relativity theory sets the speed of light as the upper bound for such speeds. Again, early theories of combustion hypothesize that some substances contain phlogiston, but this is a distortion with respect to the composition of such substances because there is no phlogiston.

Idealizations often are treated as distortions. Ronald Laymon writes, “The most natural attitude to take towards idealizations . . . is to assume that their use introduces distortion or bias into the . . . analysis” ([65], p. 354). According to Margaret Morrison, an idealization is “a characterisation of a system or entity where its properties are deliberately distorted in a way that makes them incapable of accurately describing the physical world” ([85], p. 38 fn. 1). Likewise, Nancy Cartwright takes an idealization to be a mental rearrangement or replacement of inconvenient features or specific properties of a concrete object with factors “which are easier to think about, or with which it is easier to calculate” ([17], p. 187).

If idealizations are distortions, then an idealization replaces one description of a system with a simpler description that attributes to that system at least one feature that the system does not have. The resultant description, a distorted description, either attributes an incorrect magnitude to some property of the system of interest or qualitatively distorts some property of that system. Hence, if idealizations are distortions, an idealized description of a system is an incorrect description of that system. Note that this interpretation of idealizations does not entail that every distortion is an idealization; only those distortions that perform the appropriate functions qualify as idealizations.

The interpretation of idealizations as distortions provides a semantic meaning to mathematical syntax that satisfies the characteristic functions of idealizations. In order to illustrate this interpretation of idealizations, consider two systems in which idealizations are treated as distortions.

As a distortion, the **simple pendulum** is a pendulum subject to no friction or other non-gravitational forces; it has an extension of zero; its string is rigid and has no mass; and so on. The simple pendulum lacks features that real pendula have, and it has features that real pendula lack. For instance, real pendula are subject to a non-zero amount of friction and have a finite, non-zero extension. (See Figure 1.1.) Accordingly, a description of the simple pendulum is false of every real pendulum. Nonetheless, treating real pendula as simple pendula allows complicated (hard-to-solve) equations of motion for real pendula to be replaced with equations that are simpler (easier to solve).

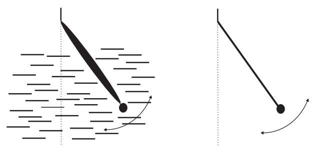

Figure 1.1: Real Pendulum vs. Simple Pendulum

Idealizations also are applied to gases. As a distortion, an **ideal gas** is a gas in which collisions between particles are elastic, in which the forces between its particles have no magnitude, etc. An ideal gas lacks features that real gases have, and it has features that real gases lack. For instance, collisions between particles of real gases are not elastic and intermolecular forces in real gases have finite, non-zero values. Accordingly, a description of an ideal gas is false of every real gas. Nonetheless, treating real gases as ideal gases allows complicated (hard-to-solve) equations of state for real gases to be replaced with equations that are simpler (easier to solve).

1.1.3 Accommodating Idealized Explanations

If idealizations are distortions then, according to the representative accounts of scientific explanation, there are no idealized explanations. Consider, first, Hempel's DN model. Hempel requires that every member of a putative explanans be true in order for it to be explanatory. If an explanation is idealized, then at least one member of its explanans is an idealization. If idealizations are distortions, this member is false. Hence, putative explanations that involve distorting idealizations are not explanatory according to Hempel's account.

Hempel attempts to avoid this conclusion by appealing to the notion of a proviso (see [44]). A *proviso* is a statement that the conditions under which an idealized law is true obtain. Hempel's strategy with provisos is two-fold. First, for any law that occurs as an element of a putative explanans, if the law is true only of idealized versions of real systems, it is to be replaced by a slightly different law. The consequent of this new law is to be the consequent of the original law. The antecedent is to be the antecedent of the original law conjoined with a proviso, which expresses the conditions under which the original law is true. Secondly, the putative explanans is to be supplemented with this proviso. This strategy prevents idealized laws from being false and allows one to infer that the idealized law is true when the conditions stated in the proviso obtain. The strategy is supposed to avoid violations of the factual correctness requirement.

Consider an example given by Hempel that illustrates the use of provisos. Suppose that β is a metal bar to which iron filings are clinging, and the explanandum is that, when β is broken into two shorter bars and the shorter bars are suspended close to each other at some distance from the ground, the bars orient themselves so as to fall into a straight line. (See Figure 1.2.) From the theory of magnetism, it is possible

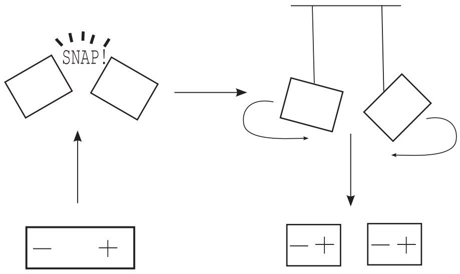

Figure 1.2: Orientation of Broken Magnet

to deduce the law that if a magnet is broken into two bars, then both resultant bars are magnets whose poles attract or repel each other. This law can be used to deduce the explanandum. Yet, as Hempel notes, the law about magnets is false: if the bar β is broken at high temperatures, it becomes demagnetized. Hence, the explanation requires a proviso to the effect that the bar is not broken at high temperatures. This proviso is added to the explanans, and the law about magnets is replaced with a law stating that if a magnet is broken into two bars but not broken at high temperatures, then both resultant bars are magnets. Hempel takes this to allow for an explanation of the explanandum, when in fact β is not broken at high temperatures.

Hempel's appeal to provisos appears to allow idealized laws to be explanatory in some cases, because sometimes real systems satisfy the conditions set forth in the proviso. This appears to avoid the conclusion that putative explanations involving idealizations are not explanatory if idealizations are distortions. But the appearance is deceptive. When a real system satisfies the conditions set forth in the proviso, the idealizations that characterize the idealized law governing the system are not distortions; instead, they are true of the system, and the law that governs the system is not an *idealized* law. When a real system does not satisfy the conditions set forth in the proviso, the proviso is false. When a putative explanans contains a false proviso, it fails to be explanatory because the factual correctness requirement is violated. Hence, the conclusion previously drawn remains valid: any explanans that contains an idealization is not explanatory according to Hempel's account, if idealizations are distortions of real systems.

Like Hempel's account, Salmon's causal account does not permit idealized explanations if idealizations are distortions. Any idealized (distorted) version of a real system lacks some detail about the causal nexus of the real system and, moreover, contains details that are false of the causal nexus for the real system. Hence, every idealized version of a real system fails to exhibit the real system as it is embedded in its causal nexus. Thus, according to Salmon's account, idealized versions of real systems are not causally explanatory of those systems.

In an attempt to avoid this result, one might weaken Salmon's requirement on causal explanations. Instead of requiring an explanation to exhibit a phenomenon in its entire causal nexus, one might require only that an explanation exhibit a phenomenon in some relevant portion of its causal nexus. Paul Humphreys suggests this approach ([48], pp. 287-288). Humphreys distinguishes between complete and true causal explanations. A complete causal explanation of an event cites all and only the causes of the event; a true causal explanation cites causes that are causally relevant to the event. (Humphreys defines a causal factor X as causally relevant to an event e if a change in X invariably results in a change in e (p. 294).) While every complete causal explanation is a true causal explanation, some true causal explanations are partial rather than complete.

The distinction between complete and true explanations is designed to allow for causal explanations that cite some, but not all, of the causes of an event. Yet the distinction does not circumvent the conclusion that idealized versions of real systems are not causally explanatory of those systems if idealizations are distortions. As a distortion, an idealization of a causal interaction replaces the actual interaction with either a non-actual interaction or no interaction at all. In either case, the resultant idealized causal nexus is not a true-but-incomplete version of an actual causal nexus. Rather, it is an incorrect version of the actual causal nexus, because it is a distortion of that nexus. A similar point holds for idealized causal processes. Given the requirement that every member of an explanans must be factually correct, every idealized (distorted) version of an actual causal nexus is factually incorrect. Accordingly, even if causal explanations can be partial, idealized (distorted) versions of actual systems are not causally explanatory of those systems.

Finally, consider Kitcher's account of explanation as unification. Suppose that idealizations are distortions. Then there is an idealized explanation, according to Kitcher's account, just in case the belief corpus of science in the limit of its rational development contains an argument that instantiates an explanatory argument pattern and this argument contains, as one of its premises, an idealization. Since, by hypothesis, idealizations are distortions, they are false. But an argument instantiates an explanatory argument pattern only if all of its premises belong to the belief corpus of scientific practice in the limit of its rational development. Since Kitcher takes the members of this corpus to be true, no idealization is a premise in any explanatory argument pattern. Hence, Kitcher's account disallows idealized explanations if idealizations are distortions.

One might attempt to avoid this conclusion by invoking Kitcher's distinction between correct explanations and acceptable ones. Correct explanations are genuinely explanatory while acceptable explanations are explanatory so far as the present scientific community can tell. Even though Kitcher's account disallows correct idealized explanations, it seems to permit acceptable idealized explanations, because the premises of an explanation that is acceptable relative to the present scientific community need only be endorsed as true at present – and this is compatible with those premises not being endorsed as true in the limit of the rational development of scientific practice.

Although the distinction between correct and acceptable explanations allows idealized explanations in principle (albeit acceptable rather than correct ones), for the most part it fails to allow them in practice. Any idealized explanation acceptable to the present scientific community is an argument in its belief corpus instantiating an acceptable explanatory argument pattern and containing an idealization as one of its premises. Since most idealizations that appear in putative explanations are known to be idealizations, most idealizations are not only false but also known to be false, if idealizations are distortions. And since an argument is in the belief corpus of a scientific community only if all of its premises are endorsed as true by that community, most idealizations are not part of the belief corpus of the present scientific community. Hence, for the most part, Kitcher's account disallows acceptable idealized explanations. (See Chapter Two for a modification of Kitcher's account that accommodates (correct) idealized explanations by allowing falsehoods to be explanatory.)

The representative accounts of scientific explanation entail that no explanations contain any distorting idealizations. This result does not depend upon the details that differentiate these accounts from one another. Rather, the result follows because the accounts accept an adequacy condition on explanation, according to which any scientific account that involves appeal to a falsehood is not explanatory. An idealized account is a putative explanation that involves appeal to at least one idealization. If idealizations are distortions, such accounts involve appeal to at least one falsehood. Hence, given this adequacy condition on explanation, no such account is explanatory.

Any account of explanation that treats idealizations as distortions and accepts a factual correctness adequacy condition on explanation entails that there are no idealized explanations. This result is worth emphasizing, in part because it seems not to have been noticed. The result does not show, however, that there are no idealized explanations. It merely provides a constraint for any philosophical account of explanation that accommodates idealized explanation. *Any philosophical account of scientific explanation that accommodates idealized explanation must either abandon the interpretation of idealizations as distortions or allow some falsehoods to be explanatory.* This result provides a framework for discussing how the statistical mechanical accounts of phase transitions and irreversibility can be explanatory despite being (ineliminably) idealized.

1.2 Clarifications

The preceding discussion might raise the following worries: (1) that the project of this dissertation is hopeless because there are no idealized explanations; (2) that idealizations should be characterized according to syntactic rather than operational criteria; (3) that it is not possible to interpret idealizations as anything other than distortions. Before considering these objections, I shall enter one remark about the kind of explanandum that is the concern of this dissertation.

There are two types of phenomenon that one might want to explain. (For a similar distinction, see [4], [6].) A *type-1 phenomenon* is an individual event or class of events. A *type-2 phenomenon* is a multiply realized pattern among some class of events. Typical type-1 phenomena include the irreversible behavior of a particular gas, or the decrease of a particular pendulum's period upon a decrease in the distance between its pivot and center of mass, or the buckling of a particular strut under a sufficiently heavy load. Typical type-2 phenomena include the shared pattern of spacing between rainbow fringes, or the general buckling of struts under sufficiently heavy loads.

A statistical mechanical account of phase transitions is an account of why a phase transition occurs in some particular system, rather than an account of why diverse systems share certain commonalities when undergoing phase transitions. And a statistical mechanical account of irreversible behavior is an account of why a particular system exhibits irreversible behavior, rather than an account of why diverse kinds of systems share certain commonalities when behaving irreversibly. Since the accounts of phase transitions and irreversibility concern type-1 phenomena, this dissertation focuses exclusively on explanations of this type of phenomena.

1.2.1 The Possibility of Idealized Explanation

One might object to the project undertaken here on the grounds that no scientific explanations are idealized. That is, one might be content to accept the conclusion that there are no idealized explanations, insisting that idealizations are distortions and only truths can be explanatory. There are two reasons to resist this attitude.

First, if no explanations are idealized, then our best sciences provide very few explanations. Most scientific accounts are idealized in some way or other. For instance, many journal articles in physics undertake to advance our scientific knowledge of some narrow domain of natural phenomena, in virtue of explaining those phenomena; and the results in these articles nearly always appeal to idealizations. Any philosophical account of scientific explanation that forbids idealized explanations thereby fails to make sense of much of scientific practice. It is counter-intuitive to hold that our best sciences furnish very few explanations. Nor is it satisfactory to treat these results as mere stop-gaps on the way to results with explanatory power, for this fails to explain why the results are taken to be a contribution to our shared scientific knowledge, and thereby fails to explain why they are taken to have epistemic, rather than merely heuristic, value. If science provides us with any explanations at all, there must be something about idealized descriptions that has explanatory power for physical systems. The philosophical task is to understand how scientific explanations can be explanatory despite being idealized, not to deny that explanations can be idealized.

Second, there are scientific accounts of phenomena that appear to be explanatory despite being idealized. Consider the rough, qualitative proportionality between a pendulum's period and the distance from its pivot to center of mass. If this distance were to increase, the period would increase; and if the distance were to decrease, the period would decrease. (This proportionality holds for most typical pendula, but there are exceptions.) Galileo took advantage of this proportionality in designing the *pulsilogium*, the first instrument to objectively measure pulse speed (see [77], pp. 88 - 90). The *pulsilogium* is a pendulum attached to a movable peg that runs the length of a scaled board. The peg is connected to the string of the pendulum: moving the peg alters the length of the pendulum. The pendulum hangs perpendicular to the scale. (See Figure 1.3: the top picture shows the *pulsilogium* from above, with the hole indicating where the pendulum hangs down through the board and the star indicating the movable peg; the bottom picture shows the *pulsilogium* from the side.)

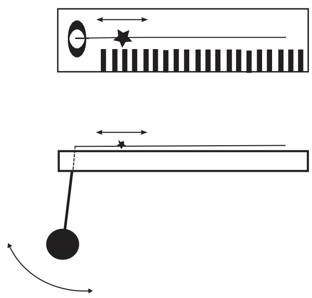

Figure 1.3: *Pulsilogium*

Given the rough proportionality between a pendulum's period and the distance between its pivot and center of mass, one can expect to alter the period of the pendulum by sliding the peg across the scale, thereby altering the length of the pendulum string. Different periods correspond to different string lengths, which correspond to different markings on the scaled board. Thus, one can measure a patient's pulse speed by sliding the movable peg until the pendulum oscillates in rhythm with the patient's pulse. This is a remarkable phenomenon, and one might very well wonder why the period of a pendulum is roughly proportional (in a qualitative sense) to the distance between its pivot and center of mass.

One account of this proportionality appeals to the simple pendulum. The simple pendulum is an idealized version of a real pendulum. A real pendulum has several characteristic properties. All of its components have a non-zero mass and extension, and imperfect rigidity. There is a point at which the pendulum connects to some support structure; this point is the pivot about which the pendulum oscillates; and the support structure itself might oscillate. There might be friction at the pivot point, which tends to impede the oscillations of the pendulum; there might be a mechanism that drives the pendulum, tending to enhance its oscillations. The pendulum also swings in some medium, such as air, oil, or water; this medium further impedes the pendulum's oscillations.

The simple pendulum is quite unlike any real pendulum. The simple pendulum is an extensionless point at the end of a massless rigid string. The point contains the pendulum's entire mass. The point at which the simple pendulum connects to its support structure is fixed, and there is no friction as the pendulum pivots about this point. There is also no resistance from the surrounding medium; it is as if the simple pendulum oscillates in a vacuum. Nor is there any mechanism that enhances the simple pendulum's oscillations; only gravity affects its motion. The simple pendulum is an idealized version of real pendula, because it idealizes many properties of real pendula. A more complete list of properties of real pendula that the simple pendulum idealizes includes: the shape, rigidity, and mass distribution of real pendula, as well as variations therein; non-gravitational forces such as air resistance and friction between any parts of real pendulum systems; driving forces; variations in the temperature, mass or length of real pendula; variations in the magnitude of gravity with altitude and variations in the direction of gravity with location on an irregularly shaped earth; and movement of the support structure.

In virtue of these idealizations, the period of the simple pendulum is roughly proportional to the length of its string. The simple pendulum can be used to predict that if the distance between a pendulum's pivot and center of mass were to increase (decrease), then the period of the pendulum would increase (decrease). The simple pendulum has all the properties that are relevant to this proportionality between pendulum period and pendulum length. The equation that governs the motion of the simple pendulum is also law-like, since it is derivable from basic principles of Newtonian mechanics (along with appeal to several idealizations, of course); and the argument pattern used to derive the equation for the period of the simple pendulum is a common argument pattern within Newtonian mechanics, suggesting that the pattern is, in some sense, a unifying pattern. Reasons like these support the claim that the explanans that explains the rough, qualitative proportionality between a simple pendulum's period and the length of its string also explains this proportionality as it exists in real pendula.

A second phenomenon amenable to idealized explanation is the familiar fact that a faint fog forms around the opening of carbonated drinks when they are first opened. This is most apparent in champagne, but it also appears in soda. (See Figure 1.4.) This phenomenon can be explained, in part, by appealing to the ideal gas law. The ideal gas law governs ideal gases, gases that are idealized versions of real gases. Unlike real gases, an ideal gas is one in which every collision between its components is perfectly elastic and in which there are no attractive forces between any of the

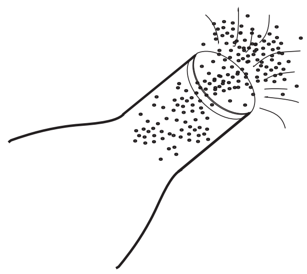

Figure 1.4: "Soda Fog"

components. Using kinetic theory, it is possible to derive the well-known law that governs ideal gases: $pV = nRT$, where p is the absolute pressure of the gas, V is its volume, T is its absolute temperature, n is the number of moles of the gas, and R is the universal gas constant. (One mole of gas contains 6.02×10^{23} particles of the gas.)

The ideal gas law relates the pressure, volume, and temperature of a gas, and this relation is a central component for an explanation of the phenomenon of "soda fog". An unopened container of carbonated beverage contains a carbonated fluid topped by a gas of water vapor and carbon dioxide (or whatever gas is used for carbonation). The pressure of this gas is greater than the atmospheric pressure, so that the pressure of the gas decreases towards the atmospheric pressure when the container is opened. Treating this gas as an ideal gas, the ideal gas law shows that, as the pressure of the gas decreases, its volume increases; that is, when the container is opened, the gas starts to expand beyond the opening of the container.

One consequence of the ideal gas law is that an ideal gas does positive work (in the technical sense of 'work') as it expands. The only source of this energy is the internal energy of the gas. Hence, according to the first law of thermodynamics, the internal energy of the gas must decrease as the gas expands. This change in internal energy is proportional to a change in temperature, for ideal gases. Thus, as the gas expands beyond the opening of the container, it undergoes a temperature decrease. This temperature decrease results in condensation of some water vapor in the gas, and this condensed vapor is the "soda fog".

Despite the idealizations that characterize an ideal gas, the ideal gas law can be used to predict that a faint fog forms around the opening of carbonated drinks when they are first opened. An ideal gas has all the properties that are relevant to the formation of this "soda fog". The ideal gas law is law-like, since it is derivable from basic principles of kinetic theory (by appeal to several idealizations). The argument pattern used to derive the ideal gas law is a common argument pattern within kinetic theory, suggesting that the pattern is, in some sense, a unifying pattern. Reasons like these support the claim that the explanans for the explanation of the formation of "soda fog" in ideal gases also explains the formation of "soda fog" in real gases.

1.2.2 Syntactic Characterizations of Idealizations

Regardless of whether there are idealized explanations, one might object that idealizations have been improperly characterized. The characterization of idealizations adopted in this chapter only appeals to the operational role of idealizations – they replace one description of a system with a simpler description. There are other characterizations of idealizations that appeal to strictly syntactic marks of idealizations without referring to the operational role of that syntax. For instance, Leszek Nowak takes the distinctive mark of idealizations to be their mathematical form, and he takes this form to be one in which the magnitude of some property of a system is set equal to zero ([88]). So, according to Nowak, if x represents a property of a system, such as its mass, radius, charge, etc., an idealization of x is a sentence to the effect that $x = 0$. Karl Popper proffers a similar characterization, prompting him to call idealization the "zero method" ([93], p. 141 fn. 2).

Nowak and Popper's approach to characterizing idealizations has the drawback that putative idealizations lacking the appropriate mathematical form are disqualified from being idealizations, despite the fact that they are mathematically equivalent to other syntax with the appropriate mathematical form. For example, a putative idealization is $g(h) = C$, where $g(h)$ represents gravity as a function of height above the Earth's surface and C is a non-zero constant (typically 9.81 m/s^2). If idealizations are distortions, this says that the force of gravity on an object is the same at all heights. This is a putative idealization, because the gravitational force on an object varies slightly as a function of the height of that object above the Earth's surface. Since this idealization does not have the form $g(h) = 0$, it does not qualify as an idealization according to the Nowak-Popper criterion. However, the syntax $g(h) = C$ is mathematically equivalent to the syntax $dg/dh = 0$, where dg/dh represents the rate of change of the gravitational force on an object with respect to its height above the Earth; given the rules of the calculus, $g(h) = C$ and $dg/dh = 0$ are derivable from each other (differentiate the former with respect to h or integrate the latter with respect to h). This latter equation qualifies as an idealization according to the Nowak-Popper criterion, even though it is mathematically equivalent to the equation $g(h) = C$, which does not qualify as an idealization according to the same criterion.

This oddity of the Nowak-Popper criterion can be avoided by using non-syntactic, functional marks to characterize idealizations. The equations $g(h) = C$ and $dg/dh = 0$

perform the same operational role and thereby produce replacement descriptions that are equally simple (in whatever sense of simplicity is appropriate). Hence, according to the operational characterization of idealizations, $g(h) = C$ is an idealization just in case $dg/dh = 0$ is. Of course, this is not to say that the Nowak-Popper criterion cannot be modified to yield the result that $g(h) = C$ also counts as an idealization. For instance, one might modify the criterion so that the distinctive mathematical form of an idealization is $x = C$, where C is a constant that is sometimes zero.

It is likely that more extensive modifications than this will be required for an adequate syntactic characterization. For instance, the criterion will need to accommodate cases in which the magnitude of some quantity is made to approach a value of infinity. An example of this kind of case is the idealization according to which the Earth's radius R is made to be infinitely large. This can be construed as idealizing the earth to be flat rather than curved, since in the limit $R \to \infty$, the curvature of the Earth vanishes.

Supposing that such a modified syntactic criterion is available, there remains the task of distinguishing between syntax that meets the modified criterion and yields a simpler description than the original description, from syntax that meets the modified criterion and does not yield a simpler description. For instance, the statement that the charge on a neutron is zero satisfies Nowak's criterion for being an idealization, but this statement is used to obtain a *correct* description of neutrons rather than a *simpler* description. *Prima facie*, there is no way to mark such a distinction by appeal to syntactic criteria alone. Moreover, many different kinds of syntax can satisfy the same function, of replacing one description with a simpler description. So rather than develop a syntactic characterization of idealizations only to append to it a non-syntactic criterion, it is more straightforward to neglect the syntactic marks of idealizations altogether. For this reason, an operational, non-syntactic characterization is preferable to the Nowak-Popper characterization and is, accordingly, the characterization of idealizations adopted for the dissertation.

1.2.3 Alternative Interpretations of Idealizations

Regardless of whether there are idealized explanations, and even if idealizations should not be characterized with purely syntactic criteria, one might object that idealizations are distortions by necessity. That is, one might insist that part of what it is for something to be an idealization is for it to be a distortion (see [88], pp. 31ff; [50], p. 175). This insistence is resisted for the following reason.

Many equations of mathematical physics appear to represent relations among properties of unobservable entities. (My use of 'represent' and its cognates is intended to be neutral regarding the success or correctness of the representation.) For instance, Coulomb's law appears to represent the force F between two charged particles as a relation between the charges of each particle, q_1 and q_2, and the distance r between the particles (k is a constant):

$$F = k\frac{|q_1||q_2}{r^2}|$$

But force and charge and the particles themselves are, in some sense, unobservable. Maxwell's equations of electromagnetism appear to represent relations between electric and magnetic fields. These fields are also, in some sense, unobservable. Most famously of all, Schrödinger's equation in quantum mechanics – $d^2\psi/dx^2 + k^2\psi = 0$ (for a particle moving in one dimension and subject to no force) – appears to represent the dynamical behavior of an unobservable wave function ψ. Within twentieth century philosophy of science, there is a history of worrying about whether these pieces of mathematical syntax in fact represent the way the world is. For instance, typical instrumentalist interpretations of quantum mechanics treat Schrödinger's equation as a mere calculation device that does not represent any "wave function". And more global versions of instrumentalism treat all such equations – indeed, all claims about unobservable entities – as mere "inference tickets" that permit inferences from old observational claims to new observational predictions.

The philosophical community's treatment of these worries about whether such syntax makes representational claims as legitimate, and the status of instrumentalism as a genuine rival to other interpretations of science, suggests the truth of the following *Instrumentalist Conditional*:

> It is possible that a piece of mathematical syntax does not represent the way the world is if it is possible to interpret that syntax as not representing the way the world is.

The Instrumentalist Conditional, supplemented with a story telling us how claims about unobservable entities can be treated as "inference tickets", entails the possibility that an instrumentalist interpretation of such claims is correct. There is no reason to suppose that the Instrumentalist Conditional applies only to equations of mathematical physics (or other claims) that appear to represent relations among properties of unobservable entities. The Instrumentalist Conditional might also apply to syntax that satisfies the functional criteria for being an idealization. If it does, and if it is possible to interpret that syntax as not representing the way the world is, then it is possible that such syntax does not represent the way the world is.

Given the absence of reasons to the contrary, it is reasonable to suppose that the Instrumentalist Conditional applies to syntax that satisfies the operational criteria for being an idealization. Moreover, it is possible to interpret this syntax as not representing the way the world is. For instance, it is possible to interpret this syntax as syntax that allows us to ignore certain features of the world (in the pursuit of certain descriptive, predictive, or explanatory aims, within certain margins of tolerable error), rather than as syntax that is a misrepresentative statement about those features. Chapter Four further elaborates and substantiates this claim. Hence, it is possible for syntax to satisfy the functional criteria for being an idealization without representing the way the world is. A fortiori, it is possible that such syntax does not *incorrectly* represent the way the world is. Therefore, since it is possible for putative idealizations (pieces of syntax) to satisfy the functional roles of idealizations and for none of these putative idealizations to be distortions, it is not necessary that all (or even any) idealizations are distortions. And if some putative idealizations satisfy the functional criteria for being idealizations but happen not to be distortions, they should not be interpreted as distortions even though they are idealizations.[2]

1.3 Chapter Synopses

Having introduced the project undertaken in this dissertation, I shall provide a rough outline of the dissertation itself.

Chapter Two surveys philosophical accounts of idealized explanation that allow falsehoods to be explanatory. In particular, the chapter considers the accounts given

[2]In giving this argument, I am not endorsing an instrumentalist account of laws. My goal is to give an instrumentalist-like account of idealizations that is compatible with a non-instrumentalist account of laws and other non-idealized elements that occur in various explanans – and to do this without undermining the explanatory power of accounts that invoke idealizations.

by Ronald Laymon, Alexander Rueger and David Sharp, Philip Kitcher, and R.I.G. Hughes. The challenge for these accounts is to show how an explanans can be explanatory despite containing a falsehood, and thereby show how an idealized explanans can be explanatory despite containing a falsehood. Chapter Two presents the main details of these accounts. The chapter is largely expository, serving as a prelude to the critical discussion in the next chapter. (The chapter is also, to the best of my knowledge, the first discussion to set these accounts side by side.)

Chapter Three confronts the accounts of idealized explanation from Chapter Two with the statistical mechanical accounts of phase transitions and irreversibility. The chapter provides relevant details of these scientific accounts, as well as arguments that both accounts are ineliminably idealized – thereby substantiating one of the main theses of this dissertation. The chapter also argues that the philosophical accounts of idealized explanation surveyed in Chapter Two do not show how these scientific accounts are explanatory. Given the presumption that the scientific accounts are explanatory, it follows that the philosophical accounts from Chapter Two are inadequate. (This presumption is defended in Chapter Five.)

There are two reasons that would explain the inadequacy of the philosophical accounts from Chapter Two. Either those accounts do not correctly identify the conditions under which falsehoods can be explanatory, or no philosophical account that takes idealizations to be distortions can accommodate ineliminably idealized explanations. Chapter Three presents a paradox of ineliminable idealization, which shows that no philosophical account of idealized explanation that takes idealizations to be falsehoods can show, even in principle, how ineliminably idealized explanations are explanatory. That is, the paradox shows the existence of ineliminably idealized explanations to be incompatible with the treatment of idealizations as distortions. Given the presumption that the ineliminably idealized accounts of phase transitions and irreversible behavior are explanatory, the key conclusion of Chapter Three is that the interpretation of certain idealizations as distortions is mistaken.

Conservatively speaking, the paradox of ineliminable idealization only shows that some idealizations are not distortions, namely, the ineliminable ones that occur in the explanations of phase transitions and irreversible behavior. In the absence of an independent, principled reason to interpret some idealizations as distortions but not others, it is ad hoc to limit the conclusion of the paradox to the claim that only some idealizations are not distortions. A uniform interpretation of idealizations is preferable to a disjoint interpretation, if a uniform interpretation is possible. The aim of Chapter Four is to provide such an interpretation.

Rejecting the interpretation of idealizations as distortions, Chapter Four presents a semantic interpretation of idealizations according to which idealizations are abstractions. According to this interpretation, idealizations are not statements (declarative sentences) and, accordingly, do not incorrectly describe the way the world is; rather, they are "inference tickets" that allow us to ignore certain features of the world without thereby misrepresenting it. If idealizations are distortions, then an idealized explanans contains at least one falsehood; whereas if idealizations are abstractions, then an idealized explanans is incomplete (but does not necessarily contain a falsehood).

Chapter Four also partially develops a philosophical account of idealized explanation that is appropriate to an interpretation of idealizations as abstractions. The account respects the intuition that falsehoods are not explanatory by treating idealized explanation as a kind of incomplete explanation. (However, I do not argue that falsehoods are not explanatory; I endorse that constraint for the sake of argument.) After presenting a partial account of idealized explanation, the chapter proceeds to show how, on the account proposed, the statistical mechanical accounts of phase transitions and irreversible behavior are explanatory. The content of Chapter Four satisfies the overarching goal of this dissertation.

Chapter Five defends the presumption that the statistical mechanical accounts of phase transitions and irreversible behavior are explanatory. Given the philosophical account of idealized explanation developed in Chapter Four, this presumption entails the striking conclusion that sometimes the correct description of a system requires the omission of details about the system. Chapter Five addresses two kinds of argument against the presumption that some explanations are ineliminably idealized. In the course of defending the presumption, I also make a case for considering phase transitions and irreversible behavior to be emergent properties of real systems.

Chapter Six provides an independent motivation for abandoning the interpretation of idealizations as distortions in favor of the interpretation of idealizations as abstractions. The motivation comes from cases in which there are two idealized hypotheses about a system, both of which are explanatory but cannot be accepted simultaneously as characterizations of the system, owing to their incompatibility with each other. Ordinarily inference to the best explanation provides a method for deciding which of the hypotheses should be taken as characterizing the system, thereby connecting explanation and ontology. Yet inference to the best explanation is not a cogent form of inference if idealizations are falsehoods. For, under such an interpretation, idealized hypotheses are false, but the conclusion of an inference to the best explanation is that the hypothesis with the most explanatory power is probably true.

Chapter Six critically discusses extant alternatives to inference to the best explanation as a method of deciding which idealized hypothesis, from a set of competing idealized hypotheses, should be taken to characterize the system of interest. There are two such accounts, one given by Lawrence Sklar, the other by Paul Teller. The chapter argues that neither of these accounts adequately characterizes the connection between idealized explanation and ontology. The account further argues that interpreting idealizations as abstractions provides an adequate characterization of this connection, in accordance with the account of idealized explanation given in Chapter Four.

Chapter Seven is the last substantive chapter of the dissertation. The chapter provides a second independent motivation for abandoning the interpretation of idealizations as distortions, based upon a problem due to Michael Shaffer, which challenges Bayesian confirmation theorists to show how at least some idealized hypotheses have at least some degree of confirmation. Shaffer argues that, in order to accomplish this task, one must either abandon Bayesianism or develop a coherent proposal for how to assign prior probabilities to counterfactual conditionals. This chapter develops a Bayesian reply to Shaffer's challenge that avoids the issue of how to assign prior probabilities to counterfactuals. The reply treats idealized hypotheses as abstract descriptions and idealizations as abstractions. It allows Bayesians to assign non-zero degrees of confirmation to idealized hypotheses and to capture the intuition that less idealized hypotheses tend to be better confirmed than their more idealized counterparts.

CHAPTER 2

EXPLANATORY FALSEHOODS

Any successful account of idealized explanation must either abandon the interpretation of idealizations as distortions or allow some falsehoods to be explanatory. Whereas the accounts of explanation surveyed in the previous chapter forbid idealized explanations, most extant accounts designed to accommodate such explanations allow some falsehoods to be explanatory, thereby allowing some distorted-because-idealized descriptions to be explanatory. Such accounts also treat idealizations as distortions; this is their motivation for providing conditions under which falsehoods can explain. This chapter presents the extant philosophical accounts of how idealized descriptions can be explanatory despite being false. (This is, to the best of my knowledge, the first discussion to set all of these accounts side by side.) A subsequent chapter argues against these accounts, and argues against the interpretation of idealizations as distortions.

There are four extant philosophical accounts of idealized explanation that, implicitly or explicitly, treat idealizations as distortions. The accounts distinguish explanatory falsehoods from non-explanatory ones on the basis of whether the idealized explanans bears an appropriate relation to the explanandum. The accounts differ on the conditions they impose upon this relation.

1. Ronald Laymon allows idealized descriptions to be explanatory if they counterfactually approximate correct descriptions.

2. Alexander Rueger and David Sharp allow idealized descriptions to be explanatory if they qualitatively approximate correct descriptions.

3. Philip Kitcher allows idealized descriptions to be explanatory if the error due to the idealized description is either negligible or unlikely to make a non-negligible difference.

4. Ronald Giere, R.I.G. Hughes, and Paul Teller allow idealized descriptions to be explanatory if the systems they describe are sufficiently similar to real systems.

This chapter presents each of these accounts in turn. The chapter is largely exegetical; criticism appears in the chapter to follow.

2.1 Counterfactual Approximation

According to Ronald Laymon, idealized descriptions are explanatory, despite their falsehood, if they counterfactually approximate correct descriptions. Whether an idealized description counterfactually approximates a correct description depends upon whether there is a modal auxiliary for the idealized sketch produced by the idealized description. Laymon, accordingly, takes an idealized explanation to have two components, an idealized sketch and a modal auxiliary.

2.1.1 Idealized Sketches and Modal Auxiliaries

The first component of an idealized explanation is an *idealized sketch*, which is a sketch of a derivation of predictions about properties of an actual system from an idealized description of that system. This derivation is a sketch, because it need not be valid. The derivation only needs to be sufficiently detailed, meaning that it only needs to give the key elements of a valid derivation. Following Hempel, the derivation only needs to give "the general outlines of what might well be developed, by gradual elaboration and supplementation, into a more closely reasoned explanatory argument" ([43], p. 424).[3] The requirement of sufficient, rather than complete, detail on the derivation is for pragmatic reasons, since a completely detailed derivation might be too lengthy or too complex for a person with limited biological and epistemic resources to provide.

The second component of an idealized explanation is, on Laymon's account, a *modal auxiliary*. A modal auxiliary is an argument demonstrating that predictions of the idealized sketch would improve in their accuracy if the idealized description were made to be more realistic ([61], p. 338; [62], pp. 157-159; [64], pp. 359-360; [63], p. 367). Laymon does not define the relation of being a more realistic description ([62], p. 158). He also holds that the determination of whether one idealized description of a system is more realistic than some other idealized description of the same system, is to be made either by appeal to (implicit) background standards or via an experiment-based bootstrapping methodology ([62], pp. 155-156; [64], p. 369).

Jeffrey Koperski notes two background standards ([59], pp. 630-631). First, one idealized description is more realistic than another idealized description of the same

[3]Peter Railton's notion of an idealized explanatory text provides an heuristic tool for thinking about what it is for a derivation to be a sketch ([97], pp. 240-246). An ideal explanatory text is a set of statements in a complete, valid derivation of an explanandum E. A statement S provides explanatory information about E if knowledge of S allows us to reconstruct or understand some portion of the ideal explanatory text for E. A sketch of the derivation given by an ideal explanatory text for E is, accordingly, a set of statements that provide enough explanatory information about E to allow the reconstruction of the entire ideal explanatory text for E.

system if it involves more accurate initial conditions, more accurate values for (measurable) independent variables. Second, one idealized description is more realistic than another idealized description of the same system if it appeals to fewer idealizing assumptions or less severe ones, where (implicit) background standards also determine the severity of idealizations. If background standards fail to determine a ranking of the relative realism of a set of idealized descriptions of the same actual system, then such a determination requires an appeal to experiment. Finding the details about how an appeal to experiment can determine the relative realism of a set of idealized descriptions remains an open project.[4]

Given a ranking of the relative realism of a set of competing idealized descriptions, a particular idealized description is made to be more realistic when it is replaced by a description that is more realistic. This explains the meaning of the antecedent in the counterfactual of a modal auxiliary. As for the meaning of the consequent of that counterfactual, the predictions of an idealized sketch improve in their accuracy if the extent to which those predictions disagree with experimental measurements or observations decreases when the idealized description in the idealized sketch is made

[4]Laymon offers an account of how this happens. Consider a simple case in which there are only two competing descriptions, I_1 and I_2. Suppose that P_1 and P_2 are the respective predictions of these descriptions. If experimental measurements or observations reveal that P_2 is more accurate than P_1 with respect to some property of the system under investigation, then the bootstrapping methodology suggests that one consider I_2 to be more realistic than I_1, at least with respect to that property. The warrant for the inference from the better predictive accuracy of I_2 to the higher (relative) realism of I_2 is the assumption that there must be something about I_2 that leads to better predictions than I_1. Laymon calls this something the higher (relative) realism of I_2. Given that I_2 is more realistic than I_1, there is an expectation (but not a guarantee) that the predictions of I_2 will be more accurate than the predictions of I_1 in relevantly similar situations ([64], p. 369). The experimental determination of the relative realism of more than two idealized descriptions of the same system proceeds in a manner similar to the experimental determination of the relative realism of only two descriptions. The basic idea is that the ranking of the relative realism of a set of competing idealized descriptions corresponds to the ranking of the accuracy of the predictions of those descriptions. (Of course, rankings of predictive accuracy are sometimes controversial; this is an issue that a Laymon-inspired account must address.)

to be more realistic ([62], p. 164). Moreover, there are several ways in which an argument can demonstrate the improvability of the predictions of an idealized sketch (see [61], pp. 342-345; [62], pp. 160-161). These ways correspond to various kinds of modal auxiliaries. Without noting the details of these different ways, note that their existence shows that a good argument for the improvability of an idealized sketch does not require actually making its component idealized description more realistic and deriving predictions therefrom. The improvability only needs to be shown to be possible. Possibility can be shown, for example, by showing that the case at hand is similar to past or paradigm cases of successful improvability ([61], p. 343 fn4).

2.1.2 Illustration

An application of Laymon's account to a specific case illustrates its details. A paradigm example of an idealized description that Laymon takes to be explanatory is the description of a real pendulum as a simple pendulum. A simple pendulum is subject to no friction or other non-gravitational forces; it has an extension of zero; its string is rigid and has no mass; and so on. According to Laymon, explanations, of real pendula, that involve the description of the simple pendulum have two components.

The first component is an idealized sketch. Consider, for example, an explanation of why a particular real pendulum happens to be at a certain location after a certain amount of time spent oscillating. The idealized sketch of this explanation includes the description of the simple pendulum; a derivation, from this description and Newton's laws of motion, of the equation that describes the motion of the simple pendulum over time; a solution of that equation in terms of pendulum position; and values for various properties of the real pendulum, such as its mass and initial displacement angle. This sketch yields predictions about the location of the real pendulum at different times of its motion, predictions that probably disagree with experimental measurements of the pendulum's location.

The second component of an explanation, of real pendula, that involves the description of the simple pendulum is a modal auxiliary. For the example from the previous paragraph, the modal auxiliary would be an argument demonstrating that a more realistic description of the real pendulum would improve on the accuracy of predictions from the idealized sketch. For example, one could obtain more precise values for the mass or initial displacement angle of the pendulum, or one could make an argument that predictions would improve if one were to take into account effects due to pivot friction or resistance of the medium in which the pendulum oscillates.[5] These alterations would provide a more realistic description of the real pendulum than does the original description of the simple pendulum, according to the background standards for descriptive realism that Koperski suggests. If it is possible to make an argument that satisfies the conditions for being a modal auxiliary, then the idealized sketch and modal auxiliary explain why a particular real pendulum happens to be at a certain location after a certain amount of time spent oscillating.

The simple pendulum illustrates the general features of Laymon's account. If an idealized sketch obtained from an idealized description has a modal auxiliary, Laymon takes the idealized description to be getting something right; however, he does not have an account of *what* this something is. Laymon's account is, accordingly, rather indefinite about what kind of information idealized explanations provide about the world. His main ideas are that an idealized description must be on the right track if

[5] For an example of this latter sort, see [52].

it figures in an idealized sketch that has a modal auxiliary, that a description could not be on the right track if it were not tracking something about the world, and that being on the right track despite being false distinguishes explanatory falsehoods from non-explanatory ones.

2.2 Qualitative Approximation

Alexander Rueger and David Sharp (henceforth RS) offer an account of idealized explanation that, like Laymon's account, takes idealizations to be distortions. Their account differs from Laymon's, in that it is more specific about what kind of information about the world idealized explanations provide. According to RS, idealized explanations provide information about topological properties of the phase space portraits of real systems. A system can be in a number of different states. Each state can be represented by a point in a phase space (so-called because it is a space that represents the different phases, or states, of the system). As the system changes over time, it comes to be in different states, represented by different points in this space. These points trace a trajectory through this space. The phase space of a system plus a set of the system's trajectories is the phase space portrait of the system. The phase space portrait of a system shows the system's qualitative behaviors, such as its points of equilibrium.

RS take an explanation that involves an idealized description of a real system to be a description of certain topological properties of the idealized system, namely, the location and nature of its points of equilibrium, maxima, minima, attractor and repellor points, saddle points, and other critical points ([101], p. 208).[6] Roughly,

[6]Strictly speaking, critical points are properties of functions that describe a system, not properties of the system itself. I speak of critical points as properties of systems because it becomes cumbersome

these features are the points at which the qualitative behavior of the idealized system changes and the nature of the changes at those points. These topological properties of the idealized system are explanatory of the real system provided it can be shown that the real system also has such properties. RS offer two conditions which, when satisfied, show the real system to have the same topological properties as its idealized version.

2.2.1 Structural Stability

The first of RS's conditions is that the law that holds of the idealized system be structurally stable.

> If the law that holds of an idealized system is structurally stable, then the real system has the same topological properties as its idealized version.

A law is *structurally stable* if a perturbed version of the law is topologically equivalent to the law. A perturbed version of a law is the law plus a perturbation of that law. A perturbation of a law is a term $\varepsilon m(x)$, where the dimensionless quantity ε is 'small' (i.e., much less than 1). A law $f(x) = 0$ and its perturbed version $f^*(x) = f(x) + \varepsilon m(x)$ are topologically equivalent if there is an homeomorphism, or continuous "rubber sheet" deformation, that morphs the phase space portrait of the system for which $f(x) = 0$ holds into the phase space portrait of the 'perturbed' system for which $f^*(x) = 0$ holds.[7]

When a law that holds of an idealized version of a real system is structurally stable, the phase space portrait of the idealized system has the same topological

to speak of them as properties of functions that describe systems. For a technical definition of a critical point, consider the one given by Bradley and Sharp ([12], p. 219). Suppose a function f is defined at a point c and either its derivative $f'(c)$ is undefined or $f'(c) = 0$. Then c is a critical number of f and the point $P(c, f(c))$ on the graph of f is a critical point.

[7]See [1], pp. 365-366 for more details on perturbations, topological equivalence, and structural stability, including some very nice pictures.

properties as the phase space portrait of the real system. This is because the real system is taken to be some ‘perturbed’ version of its idealized version. Hence, the law for the idealized system correctly describes the topological features of the real system; in RS’s terminology, the law for the idealized system qualitatively approximates the (unknown) law that holds of the real system ([101], p. 214). And the structurally stable law for the idealized system explains the qualitative behavior of the real system, because this behavior is the same in both the real and idealized systems. The qualitative behavior is the same in both systems, because the structural stability of a law for an idealized system guarantees that the topological properties of the idealized system are not merely artifacts of the idealizations that go into the description of the system, but rather properties of the real system that is idealized.

2.2.2 Illustration

Consider an example, again about the simple pendulum, to illustrate the notion of structural stability. The law that describes the behavior of the simple pendulum is

$$\ddot{\theta} = -\frac{g}{L}\sin\theta,$$

where θ is the displacement angle of the pendulum, g is the force due to the gravitational field, and L is the distance between the pendulum’s pivot and its center of mass.[8] The phase space portrait for the simple pendulum is an ellipse, which is a closed curve. (See Figure 2.2.) Now consider a perturbation of the law for the simple pendulum, which results from adding a term to represent damping due to the surrounding medium. The law that describes the behavior of the simple pendulum

[8]This law has the same form as the law for a simple harmonic oscillator, which RS discuss on p. 209.

with linear damping is:

$$\ddot{\theta} = -\frac{g}{L}\sin\theta - b\dot{\theta},$$

where b is a non-negative constant whose value depends on the properties of the surrounding medium as well as the shape and dimensions of the pendulum and the attachment of the rod to the pivot. The phase space portrait for this law is a spiral,

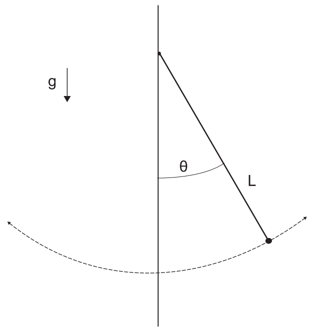

Figure 2.1: Motion of the Simple Pendulum

which is an open curve. (See Figure 2.3.) Hence, the qualitative behavior of the idealized law for the simple pendulum ($b = 0$) differs from the qualitative behavior of the law for the simple pendulum with linear damping ($b > 0$); the law for the simple pendulum is structurally unstable. Notably, and by contrast, the law for the simple pendulum with linear damping is structurally stable. This is because, according to RS, "variations in the strength of the damping force [$b\dot{\theta}$] will not qualitatively alter the dynamics" ([101], p. 209); in other words, systems subject to different amounts of damping still have spiral phase space portraits.

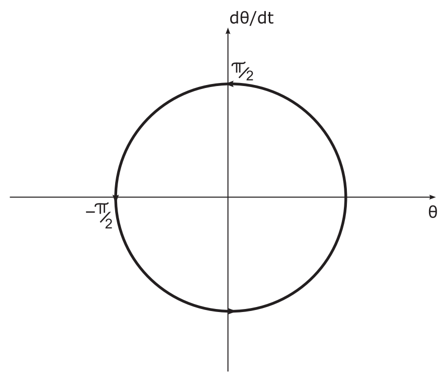

Figure 2.2: Phase Space Portrait for Undamped Simple Pendulum

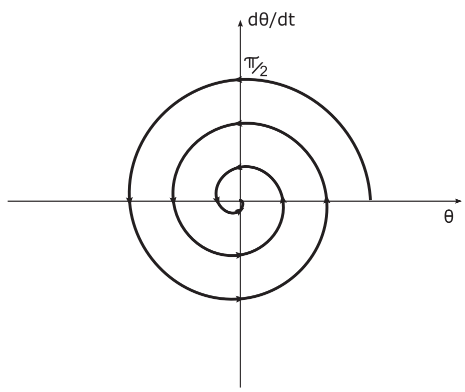

Figure 2.3: Phase Space Portrait for Damped Simple Pendulum

2.2.3 Structural Stability for Families

As the law that describes the behavior of the simple pendulum without damping shows, not all laws are structurally stable. Hence, some idealized laws (the structurally unstable ones) are not explanatory according to RS's first criterion. Nonetheless, RS deny that the structural instability of an idealized law renders that law non-explanatory of real systems. They provide a second condition which, when satisfied, shows that a real system has the same topological properties as its idealized version. This second condition is that the law family for the idealized law be structurally stable as a family.

> If the law family for the law that holds of an idealized system is structurally stable as a family, then the real system has the same topological properties as its idealized version.

Consider a law $f(x) = 0$ and parameters $u_1, \ldots, u_n$. When each parameter $u_i = 0$, $f(x) = F(x, u_1, \ldots, u_n) = F(x, 0, \ldots, 0)$. The law family for $f(x) = 0$ is the set of laws generated by allowing the parameters $u_1, \ldots, u_n$ to take on values that are small perturbations from the value zero. And a family of laws for $f(x) = 0$ is structurally stable as a family if each member of the family is topologically equivalent to the unperturbed law $f(x) = 0$ (see [101], p. 212). (Generally, laws governing chaotic systems are structurally unstable but structurally stable as families.)

For instance, the law of motion for a damped simple pendulum is a function of θ, g, L, and b, where $f(\theta, g, L, b) = 0$. Since the damped simple pendulum has a point-mass bob, the function $f(\theta, g, L, b)$ can be rewritten in terms of a function $F(\theta, g, L, b, r)$ that includes a term for the radius of the pendulum bob, viz., r. And since $r = 0$ for the damped simple pendulum, $f(\theta, g, L, b) = F(\theta, g, L, b, 0)$. Apart from the law $F(\theta, g, L, b, 0) = 0$, each member i of the law family for $f(\theta, g, L, b) = 0$

is an equation $F_i(\theta, g, L, b, r_i) = 0$ in which the parameter r_i has a specific (and small) non-zero value. This family of laws is structurally stable as a family, because each member law describes a system with behavior that is qualitatively the same as the behavior of the systems described by other family members (all such systems have a spiral phase space portrait).

Generally, if the law family for a law $f(x) = 0$ is structurally stable as a family, then either (a) every member of the family is structurally stable or (b) each member of the family is structurally unstable in the same way. (Two laws are structurally unstable in the same way if every perturbation of one law produces the same qualitative deformations of phase space portraits as the same perturbations of the other law.) If each member of a law family is structurally stable, then the law for the idealized version of the real system is structurally stable. This entails that the law is explanatory of the real system, via RS's first condition. If every member of the family is structurally unstable in the same way, then the idealized law is structurally unstable, and RS's first condition does not apply. Since, however, every other family member of this law is structurally unstable in the same way, the structural instability of the idealized law is not an artifact of the idealized description that leads to the law for the idealized system. Rather, the structural stability of the family as a family reveals that the real system itself is structurally unstable, since the (unknown) law for the real system is one of the family members of the law for the idealized system. Hence, the phase space portrait of the idealized system has the same topological properties as the phase space portrait of the real system. Consequently, the idealized system explains the qualitative behavior of the real system.

2.3 Negligibility

Acknowledging that his original account of explanation rules out the possibility of idealized explanation, Philip Kitcher offers an account of idealized explanation that abandons the requirement that every premise of every explanatory argument be a member of the belief corpus of scientific practice in the limit of its rational development (He also abandons the requirement that every premise of every presently-acceptable explanation be a member of the present belief corpus of the scientific community.) Kitcher thereby rejects the assumption that falsehoods and non-facts are not explanatory ([57], pp. 434, 452-454).

2.3.1 Prologues and Epilogues

Kitcher accommodates idealizations within his account of explanation as unification by adding a prologue and an epilogue to such explanations. (This is my terminology, not Kitcher's.) The *prologue* replaces a description of the system of interest that is accepted as correct with an idealized description of that system: "When we explain the behavior of actual objects, the first step is always to achieve an idealized description of those objects" ([57], p. 453). This replacement is justified by showing that the idealized features either make negligible differences to the phenomenon of interest or have a very low probability of making non-negligible differences. An idealized version of the phenomenon of interest is then derived through an appeal to explanatory argument patterns. The result of this derivation is an explanation of the idealized version of the phenomenon of interest. The actual phenomenon of interest is explained by an *epilogue*, which shows the ways in which the actual system differs from its idealized version and the extent to which these differences result in deviations from the explanation of the idealized version of the phenomenon of interest.

2.3.2 Illustration

To illustrate Kitcher's account, and to switch examples, consider a soldier who is firing bullets from a musket towards a target 90 meters away. The soldier can manipulate the distance the bullets travel by adjusting the inclination of the musket barrel upon firing, or by modifying the initial velocity of the bullet by adjusting the powder charge. Suppose the soldier becomes curious about why his bullets reliably traverse the 90 meters to the target whenever he fires bullets with a launch angle of 60 degrees and an initial velocity of 44.7 meters per second. (The soldier is not asking why the bullets hit the target; the question is why they travel at least 90 meters, quite apart from where they end up upon completing their journey.)

One way to answer the soldier's request for an explanation is to figure out some sort of equation that describes the motion of projectiles. The world being a complicated place, some idealization will go into obtaining an equation of motion for simple projectiles. If one assumes that the only properties relevant to the distance R the bullets travel (to a point as high as the end of the barrel) when fired from a musket are the powder charge (or, equivalently, the initial velocity v_0 of fired bullets), the

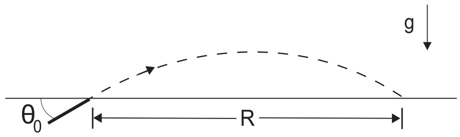

Figure 2.4: Path of a Projectile

inclination θ_0 of the barrel upon firing, and effects due to the gravitational pull of the earth, one obtains an idealized equation that relates these conditions (the initial conditions) to the distance R, namely,

$$R = \frac{v_0^2}{g} \sin 2\theta_0.$$

The derivation of this equation proceeds from premises that include the fundamental equations of Newtonian mechanics ($F = ma$) as well as the assumption that the only properties relevant to the total distance a bullet travels are its initial velocity, its angle of inclination upon firing, and gravity. The equation is idealized, because its derivation also includes appeal to several idealizations, such as the assumption that the air through which bullets move has no effect upon their motion (despite moisture or wind, say), and the assumption that irregularities on the surface of each bullet do not affect their motion.

These idealizations of bullets and the environment in which they travel constitute, on Kitcher's account, what I have called the *prologue* of an idealized explanation. They replace a correct description of bullets and their environment with an idealized description. If the idealized properties can be shown to have a low probability of making a non-negligible difference to the total distance traveled by bullets fired from the soldier's musket, then this replacement is justified. Suppose, for the sake of illustration, that this has been shown. For instance, suppose that the soldier is practicing in very thin air, high in the Rocky Mountains; that the bullets are precisely engineered to be nearly spherical, and that other situations are arranged so as to make errors due to other idealizations negligible.

Since these idealizations are justified, the equation for simple projectiles can figure in an idealized explanation of why the soldier's bullets reliably traverse the 90 meters to the target whenever they are fired (in the given environment) with a launch angle of 60 degrees and an initial velocity of 44.7 meters per second. The equation is derived in the way that many other equations are derived within Newtonian mechanics, by appealing to the fundamental equations of motion and initial conditions. If we suppose, for the sake of illustration, that Newton's equations of motion are part of the belief corpus of scientific practice in the limit of its rational development (at least for medium-sized objects), and that the derivation instantiates an explanatory argument pattern, the result is an explanation of the behavior of idealized projectiles.

The behavior of real projectiles is explained by appending an epilogue to the explanation for idealized projectiles. The epilogue catalogues the way in which the soldier's real bullets differ from idealized projectiles, and the way in which these deviations make the real bullets behave differently than their idealized counterparts. For instance, the soldier's bullets travel through a medium rather than a vacuum, and this medium tends to impede their forward progress. The soldier's bullets, unlike idealized projectiles, have slight surface imperfections and are not perfectly spherical; these geometric flaws tend to lessen the distance the bullets travel. If comments of this kind are given for every idealization mentioned in the prologue, then, since the idealized equation for projectiles predicts a total travel distance of 177 meters for a bullet fired at an initial inclination of 60 degrees and initial velocity of 44.7 meters per second, that equation explains why bullets reliably traverse at least 90 meters whenever they are fired (in the given environment) with a launch angle of 60 degrees and an initial velocity of 44.7 meters per second.

2.4 Sufficient Similarity

All of the philosophical accounts of explanation discussed so far provide precise conditions for counting something as an explanation. The general form of these accounts is something like 'If an account A of phenomenon P satisfies conditions $C_1, \ldots C_n$, then A explains P.' These philosophical accounts are precise, in the sense that it is (in principle) possible to determine, for any given presentation of an account, whether the account satisfies conditions $C_1, \ldots C_n$, without considering the various contextual and pragmatic factors surrounding the presentation of the account. This precision is absent in the philosophical account that takes scientific explanation to be metaphoric redescription.

2.4.1 Metaphoric Models

R.I.G. Hughes is representative of those who take scientific explanations to be metaphoric redescriptions. Ronald Giere and Paul Teller propose similar accounts, but Hughes's presentation is the most concise. According to Hughes,

> We explain some feature X of the world by displaying a model M of part of the world [or of the world as we describe it], and demonstrating [via proof from the model's theoretical definition] that there is a feature Y of the model that corresponds to X, and is not explicit in the definition of M ([47], p. 146).

The theoretical definition specifies a model or class of models. The subject of a model is that part of the world to which the model or theory applies (p. 137). Many kinds of models may be used in theoretical explanations; but all models should be thought of as representations (pp. 146, 147). For a model to represent is for the model to be able to "stand in" for its subject in analogous circumstances, in varying respects and to varying degrees (p. 138).

According to this kind of account, models are metaphors. A model is "a non-literal description of a primary system [the subject of the model] in terms of a secondary system" (p. 143). According to Hughes, the cognitive function of scientific models, and of metaphors generally, is to provide greater awareness of the primary system by seeing it in the terms, associations, and framework of the secondary system. Scientific models provide this secondary system, and they provide understanding and explanation of that subject through their redescription of the primary system or subject of the model (pp. 143-144). Models offer representations of the world that are adequate in varying respects and to varying degrees. We understand the world by representing it with models supplied by our theories, we display this understanding by "the ease with which we follow explanations presented in terms of the theory," and we increase our understanding of phenomena by becoming more aware of the resources of the models we use to represent them (pp. 148, 149).

Although all explanations are, according to this kind of account, metaphoric models, not all metaphoric models are explanatory. Being able to redescribe a primary system metaphorically in terms of some secondary system is necessary, but not sufficient, for having an explanation of the primary system. To be sufficient, the secondary system must be similar to the primary system in relevant respects, and the similarity between each relevant respect must be of a certain degree. On the issue of which respects and degrees are relevant in order for a metaphoric redescription to be an explanation, Teller holds that "[n]o general account is needed precisely because it is the specifics of any case at hand which provide the basis for saying what counts as relevant similarity" ([116], p. 401). Moreover, "similarity involves both agreement and difference of properties, and only the needs of the case at hand will determine whether the agreement is sufficient and the differences tolerable in view of those needs" (p. 402). Hughes and Giere share this view that various contextual and pragmatic factors surrounding the presentation of a metaphoric redescription of a phenomenon determine whether that redescription is explanatory.

The account of scientific explanation as metaphoric redescription readily provides an account of idealized explanations.[9] Idealized descriptions are explanatory just in case the idealized models they describe are sufficiently similar to real systems. Pragmatic and contextual factors determine whether an idealized model is sufficiently similar to a real system. These factors vary on a case-by-case basis. This attitude towards the conditions that suffice for an idealized model to be explanatory makes the account of idealized explanation as metaphoric redescription a vague account. Idealized descriptions, although metaphoric rather than literal descriptions of the world, are correct descriptions of the world in certain context-relative respects and to certain context-relative degrees; as such, they provide context-sensitive knowledge of the structure of the world.

2.4.2 A Worry about Vagueness

The vagueness in the criteria for similarity has prompted Lawrence Sklar to argue that the account of explanation as metaphoric redescription is unsatisfactory as an account of *idealized* explanation. For the account does not show how idealized descriptions, understood as metaphoric redescriptions of the real world, provide knowledge of the world even though they are literally false. The most the account of explanations as metaphors can say is that if an idealized model is sufficiently similar

[9]This is not to say that all of those who propose this kind of account of idealized explanation treat idealizations as distortions; for instance, Giere holds that idealizations are abstractions (see [38], p. 78).

to the real world, then it explains the real world. The account leaves context to determine whether any idealized model is sufficiently similar to the real world; nothing more can be said apart from looking at the specific purpose and situation for which a given model is used. Hence, as Sklar notes ([109], p. 42), the account of explanation as metaphoric redescription embeds into the notion of similarity an answer to the question of whether a particular idealized model is explanatory – or, in Hughes's terminology, embeds such an answer into the notion of "standing in." Since this notion is not well-understood, Sklar concludes that the account is unsatisfactory.[10]

In response to Sklar's criticism, advocates of the account of explanation as metaphoric redescription might offer a Lewisian rejoinder (see [71], pp. 91-95). They might claim that the charge of not being well-understood is ambiguous; the charge could be that similarity is ill-understood, or it could be that similarity is vague. Ill-understood notions are, of course, to be eschewed, but similarity is not ill-understood. Rather, similarity is vague. Since what counts as an explanation is also vague, varying as it does with innumerable contextual and pragmatic factors, similarity is just the kind of notion to use in an analysis of explanation. Hence, no matter how mysterious the factors that determine relevant similarity might be, analyzing explanation in terms of similarity reduces two mysteries to one; the limited vagueness of similarity nicely accounts for the limited vagueness of explanation. (Moreover, since this is a Lewisian defense, one might also point out the theoretical utility of the notion of similarity in Lewis's analysis of counterfactuals, thereby bolstering the number of mysteries that can be eliminated by appealing to this notion.)

[10]In a different context, F.A. Muller makes a similar point about the appeal to context: "Uttering '*Context!*' seems a *deus ex machina* of the past fifty years of philosophy. Without further explication and clarification, such utterances are more like performing acts of philosophical magic than propounding a philosophical argument. Philosophy is not sorcery." (See [86], p. 69 fn. 12.)

The Lewisian rejoinder might continue with a criticism of Sklar's preference for a precise account of explanation. Advocates of the account of explanation as metaphoric redescription might argue that the quest for a more exact or precisified notion of similarity is to be eschewed. Some important similarities involve "idiosyncratic, subtle, Gestalt properties," and it is impossible – in practice if not in principle – for an exact notion of similarity (or any exact notion, for the matter) to track all of this quirkiness. A more precise criterion might get extreme cases correct, but it would inevitably fail when applied to ordinary cases. Hence, according to the rejoinder, the quest for a precise analysis of an imprecise concept like explanation is misguided; the account of explanation as metaphoric redescription might be vague, but this is a merit rather than a shortcoming.

This Lewisian defense rests upon the assertion that the notion of similarity at work in the analysis of explanation as metaphoric redescription is vague rather than ill-understood. While the notion is vague, I am not convinced that it is not ill-understood. Although the notion of similarity figures prominently in our everyday lives, it does not follow that we understand what we are doing when we use that notion; nor does it follow that we understand why scientists make the judgments of similarity that they make; nor does it follow that there is nothing more informative to be said about explanation, and nothing to be said that applies apart from details of context. Of course, none of this amounts to a refutation of the Lewisian defense. And although I don't buy the adequacy of the defense, I am willing to grant, for the sake of argument, that the account of explanation as metaphoric redescription shows why some idealizations are explanatory.

2.4.3 Illustration

Even granting this, it is difficult to assess the adequacy of this account for any particular idealized explanation. Consider, once more, the idealized explanation of the distance the soldier's bullets tend to travel when fired with an initial velocity of 44.7 meters per second from a musket inclined to 60 degrees. Suppose the context is one of curiosity: the soldier would just like to understand why the bullets travel as far as they do. Perhaps, in this context, idealized projectiles are sufficiently similar to the soldier's actual bullets. If so, then the idealized equation for the distance projectiles travel is explanatory in this context. Suppose, however, that the context is one of wartime preparation, and the soldier needs to understand how far the bullets travel in order to make a report to a committee charged with arming a military for its invasion of an environment with very moist air and strong wind currents. Perhaps, in this context, the idealized projectiles are not sufficiently similar to the soldier's actual bullets. For the way those real bullets are affected by the medium through which they are intended to travel is dissimilar to the way in which idealized projectiles travel through a vacuum. If this is right, then in this context the idealized equation for the distance projectiles travel is not explanatory.

As this example illustrates, it is difficult to apply the account of idealized explanation as metaphoric redescription to specific cases without an account of which similarities are relevant (and of what the tolerable margins of error are for certain predicted magnitudes). For this reason, this account's treatment of specific instances of putative explanations that appeal to idealizations will not be informative enough to be interesting or satisfying. The most that can be said of any particular idealized account is that, if the dissimilarities of the idealized version of a system are not relevant to the system itself, and if the similarities are relevant, then the account is explanatory despite being idealized. More cannot be said, because there are no criteria with which to discuss whether the similarities and dissimilarities in particular cases are relevant. Hence, although this account of idealized explanation as metaphoric redescription is included in this chapter for the sake of thoroughness, it will not be further discussed or applied to examples in the remainder of the dissertation.

2.5 Conclusion

Any successful account of idealized explanation must either abandon the interpretation of idealizations as distortions or allow some falsehoods to be explanatory. Any account of explanation that allows some falsehoods to be explanatory must provide a criterion that distinguishes explanatory falsehoods from non-explanatory ones. In its most general form, this criterion requires false descriptions to bear an appropriate relation to their correct (non-idealized) counterparts in order to be explanatory. Different philosophical accounts of idealized explanation differ on what they take this appropriate relation to be. Candidate relations include counterfactual approximation, qualitative approximation, sufficient similarity, and negligible or improbable distortion.

Here is a summation of these accounts, for ease of reference:

- *Counterfactual Approximation*: An idealized description is explanatory of phenomenon P if (a) there is a sketch of the idealized description and (b) there is a modal auxiliary for that description, showing that the predictions of the idealized description as regards P would improve if the description were made to be more realistic (less idealized).

- *Qualitative Approximation*: An idealized description is explanatory of a *topological property* of a system if (a) the description is law-like and (b) either the idealized law is structurally stable or the law family for the idealized law is structurally stable as a family.

- *Negligibility*: An idealized description is explanatory of phenomenon P if (a) the error due to the idealized description of P is either negligible or unlikely to be non-negligible, (b) there is an account of the ways in which the idealized description of P differs from a correct description of P, (c) there is an account of the extent to which these differences result in corrections to the idealized explanation of P, (d) the argument in which the idealized description occurs instantiates an explanatory argument pattern.

- *Sufficient Similarity*: An idealized description is explanatory of phenomenon P if and only if the idealized description describes a system that is sufficiently similar (in context-relative respects, to context-relative degrees) to the system in which P occurs.

These requirements are formulated as sufficient conditions only (except for the last one); those who offer the accounts do not commit themselves to the necessity of their proffered conditions. It would be reasonable to assume that the conditions are also necessary; this would provide the most straightforward method of delineating explanatory falsehoods from non-explanatory ones. But I do not make this assumption.

The next chapter provides an argument that these accounts are inadequate. The first part of the argument shows that some idealized explanations are ineliminably idealized, that there are some phenomena that require, in principle, an appeal to idealization in order to be explained. The second part shows that these ineliminably idealized explanations do not satisfy any of the conditions set forth by the accounts surveyed in this chapter. The chapter ends with an argument for the claim that no account that interprets idealizations as distortions is compatible with the existence of ineliminably idealized explanations.

CHAPTER 3

AGAINST IDEALIZATIONS AS DISTORTIONS

Extant accounts of idealized explanation that allow falsehoods to be explanatory accept, at least implicitly, an interpretation of idealizations as distortions. The distinguishing characteristics of idealizations are that they replace one description of a system with a description that is, in some sense, simpler. It will be recalled from Chapter One that a distortion is something that attributes a feature to a system that the system does not have. There are many kinds of distortions, such as lying and ordinary misdescriptions. If idealizations are distortions, then idealized descriptions are incorrect descriptions. Specifically, if idealizations are distortions, then an idealization replaces one description of a system with a description that attributes to that system at least one feature the system does not have. The resultant description, an idealized description, is a false (distorted) description of the system.[11]

Philosophical accounts of idealized explanation that allow idealized descriptions to be explanatory despite their falsity presuppose that idealized descriptions are false. For if idealized descriptions are not false, then there is no need to show how they can

[11]Note that this description is not false solely in virtue of its being an idealized description, because idealized descriptions are merely descriptions obtained via appeal to syntax that satisfies certain criteria for being an idealization and these criteria do not specify the semantic role of such syntax. Rather, it is false in virtue of the interpretation that takes idealizations to be a kind of distortion and the fact that, necessarily, distorted descriptions are false. If idealizations were interpreted in some other way, it might turn out that idealized descriptions need not be false.

be explanatory despite being false. The interpretation of idealizations as distortions is the only interpretation of idealizations according to which idealized descriptions are false: if idealizations are not distortions, then there is no reason to suppose that idealized descriptions are always false. Hence, insofar as philosophical accounts of idealized explanation presuppose that idealized descriptions are false, they interpret idealizations as distortions.

The accounts of idealized explanation surveyed in Chapter Two provide various conditions which, when satisfied, show that idealized descriptions are explanatory despite being false. This suggests that those accounts treat idealizations as distortions, at least implicitly. The aim of this chapter is to show that the accounts surveyed in Chapter Two are inadequate and that the interpretation of idealizations as distortions is mistaken. Showing that it is a mistake to interpret idealizations as distortions suggests that it is a mistake to treat all idealzations as false. This results removes the necessity for accounts of idealized explanation that allow idealized descriptions to be explanatory despite being false.[12]

There is an important class of idealized explanations that do not satisfy any of the conditions set forth by the accounts surveyed in Chapter Two. These explanations are ones that are ineliminably idealized. (An idealized explanation of some phenomenon is ineliminably idealized if the only way to explain the phenomenon is to appeal to an idealization; likewise for ineliminably idealized descriptions.) The accounts of idealized explanation surveyed in Chapter Two do not show how ineliminably idealized

[12]As a matter of fact, the accounts surveyed in Chapter Two accept the interpretation of idealization as distortions; but this interpretation is not a mandatory component of the accounts. The criticism of those accounts to be given in this chapter does not depend upon construing those accounts as treating idealizations as distortions; those accounts would be inadequate even if idealizations were to be interpreted as something other than distortions.

descriptions can be explanatory despite being false. Moreover, no account of idealized explanation that takes idealizations to be distortions can show, even in principle, how ineliminably idealized descriptions are explanatory. This is due to what I call the paradox of ineliminable idealization, which shows that the existence of ineliminably idealized explanations is inconsistent with an interpretation of idealizations as distortions. (Details on this paradox are given in the final section of this chapter.)

This chapter contains the details of this argument against interpreting idealizations as distortions. The chapter presents, in reasonable detail, two cases of idealized explanation. The first is the statistical mechanical account of the occurrence of phase transitions. The second is a statistical mechanical account of the way in which non-equilibrium systems, when left to themselves, irreversibly approach equilibrium in a finite amount of time. Following each account, it is argued that the extant philosophical accounts of idealized explanation that allow some falsehoods to be explanatory do not show how these scientific accounts are explanatory. The reason for this is that the accounts from statistical mechanics are ineliminably idealized; each explanation requires, in some sense, the appeal to an idealization. This requirement gives rise to the paradox of ineliminable idealization, which shows, under very general assumptions, that the ineliminable idealizations in each account are not distortions.

3.1 Phase Transitions and the Thermodynamic Limit

At a pressure of about 15 kilobars (e.g., about 50 kilometers below the earth's surface), graphite spontaneously converts into diamond. At temperatures below 77 Kelvin, nitrogen liquifies. At about 373 Kelvin and normal atmospheric pressure, water boils. Iron and nickel become magnetized when in the presence of a magnet, and at ordinary temperatures stay magnetized when the magnet is removed (this is called, somewhat misleadingly, spontaneous magnetization). The electrical resistivity of pure mercury drops significantly at a temperature of about 4 Kelvin, and the mercury becomes a superconductor.

These examples, and others like them, demonstrate the pervasiveness of phase transitions in our experience of the world. A phase transition is said to occur when (roughly) a change in value of some system parameter (temperature, pressure, etc) results in a large change in the (qualitative) state of the system. For instance, raising the temperature of water from 272 Kelvin to 303 Kelvin, at normal atmospheric pressure, results in a phase transition: the ice melts into a liquid.

The different phases of a system exhibit qualitatively different large-scale behaviors. Liquid water, but not solid ice, takes the shape of its container; the solid phase of water holds its shape, and also happens to occupy a different volume than that same amount of water in the liquid phase. Solids, but not liquids, propagate transverse waves; liquids only propagate longitudinal waves.[13] Similarly, graphite, but not diamond, can be used for writing; and diamond, but not graphite, is not only very beautiful but also radiation-hard, which is why people are trying to use it to coat components in high-energy, radiation-producing particle colliders. (Graphite and diamond are different solid phases of carbon.) Again, as everyone knows, magnetized iron can hold pictures on metal refrigerator doors; unmagnetized iron cannot.

The changes in a parameter of a system need not be large in order for the system to undergo a phase transition. This is best illustrated with a pressure-temperature

[13] A wave is transverse if the oscillating elements of the medium in which the wave propagates are perpendicular to the direction in which the wave travels (like a wave sent along a taut rope). A wave is longitudinal if the oscillating elements of the medium in which the wave propagates are parallel to the direction in which the wave travels (like sound traveling through water).

diagram for a generic pure fluid in equilibrium: see Figure 3.1. This sort of diagram

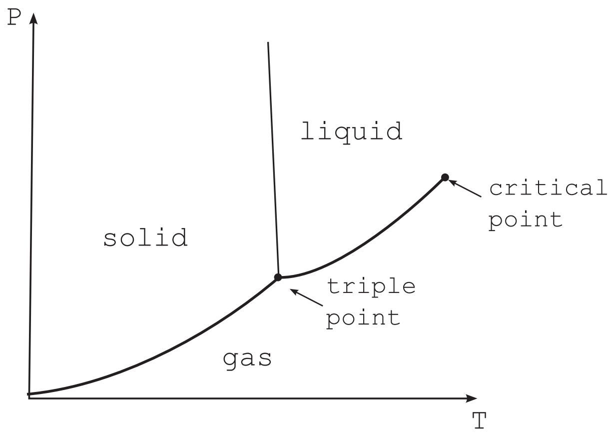

Figure 3.1: Pressure-Temperature Diagram for Pure Fluid

can be obtained through careful experimental measurements of the pressure and temperature of real fluids such as water. The lines on the diagram delineate the three phases of a fluid, when the fluids are in equilibrium states. The line between the solid and liquid phase is known as the fusion curve. A system can exist in both phases simultaneously when its temperature and pressure is at one of the points on the fusion curve; but when the system falls on either side of this curve, the system is either a liquid or else a solid. The line between the liquid and gas phases is known as the vaporization curve, and the line between the solid and gas phases is known as the sublimation curve. The point at which the fusion, vaporization, and sublimation curves intersect is known as a triple point; this is the point at which all three phases of the system can exist together simultaneously. The point at which the vaporization curve ends is known as a critical point; for temperatures greater than the temperature at this point, it is possible to transform a fluid between its liquid and gas phases without crossing the vaporization curve. (This suggests that there is no qualitative difference between the liquid and gas phases of a fluid, because it is possible to change a fluid from one phase to the other in a continuous manner.) Since the interfaces between the phases are very narrow (separated by lines), very small changes in the pressure or temperature of a system can produce very drastic changes in the system's large-scale qualitative behaviors.

Statistical mechanics is able to show how macroscopic phenomena like phase transitions result from microscopic behaviors. From a microscopic point of view, the existence of phase transitions is surprising. According to Debashish Chowdhury and Dietrich Stauffer, a

> macroscopic system is capable of exhibiting phenomena which are, apriori, not obviously expected from purely mechanical consideration of its constituents; for example, why do the molecules of a fluid condense to form a liquid at sufficiently low temperatures whereas the same molecules remain in the gaseous phase at high temperatures although in both the situations their equations of motion involve the same form of the inter-molecular interactions? ([24], p. 101)

Its account of the occurrence of phase transitions is one of the great successes of statistical mechanics.

The discussion of the statistical mechanical account of phase transitions proceeds as follows. I begin with a brief discussion of phase transitions from a thermodynamic point of view. This includes a rough discussion of the physical reasons for phase transitions, and a brief discussion of the rationale for the usual method of representing the occurrence of a phase transition. Next, I discuss the key elements in any statistical mechanical account of phase transitions, including the notion of a partition function and the ineliminability of the thermodynamic limit to any statistical mechanical description of phase transitions. Then, for purposes of illustration, I discuss the 2-dimensional Ising model of phase transitions, an exactly solvable model that predicts the occurrence of a ferromagnetic-paramagnetic phase transition in ferromagnets. This discussion of the Ising model further clarifies the ineliminability of the thermodynamic limiting idealization. Following all of this, I show that the philosophical accounts of idealized explanation (from the previous chapter) do not show how the statistical mechanical account of phase transitions is explanatory.

3.1.1 A Thermodynamic Point of View

Thermodynamics, the study of the macroscopic properties of matter, provides a qualitative understanding of the physical reasons for phase transitions. Suppose that a system is in equilibrium, with a fixed volume and number of particles and a constant temperature T. Then it is a well-known result of thermodynamics that the Helmholtz free energy of the system, F, is minimized. The Helmholtz free energy of a system is something like the energy it would take to create the system out of nothing in an environment of temperature T; this is the energy of the system itself, less the energy that can be contributed by the environment. Where E represents the internal energy of the system and S represents the system's entropy, $F = E - TS$. The energy E is due, in part, to interactions among the system's particles. These interactions tend to make the system more ordered. The entropy S, in contrast, tends to make the system less ordered. At high temperatures, F is best minimized by maximizing S, so that the TS term dominates the E term and the system remains in a disordered phase. At lower temperatures, however, ordering can occur, provided that the energy contribution to F is stronger than the contribution due to entropy. Hence, at certain temperatures, there will be a phase transition between a more ordered phase and a less ordered phase.

An example will help to illustrate this qualitative account. A metal, such as iron or nickel, may be thought of as a collection of "spins"; the metal's being magnetized may be thought of as the phase of the metal in which most of its spins are aligned in the same direction, and the metal's not being magnetized may be thought of as the phase in which the metal's spins are randomly aligned. (More on this later, in the discussion of the Ising model of ferromagnetism.) When the spins of the metal are randomly aligned, the metal is in a very disordered state; as more and more of the spins become aligned in the same direction, the metal becomes more and more ordered.

When the metal is in equilibrium at a constant temperature, its Helmholtz free energy is minimized. This minimization requires the "best" compromise between minimal energy, for which all of the spins are aligned, and maximal entropy, for which the spins are randomly aligned. (Because, in a constant-temperature environment, saying that F tends towards its minimum possible value is the same as saying that E tends towards its minimum possible value while S tends towards its highest possible value.) At low temperatures, the best way for the metal to have a minimum Helmholtz free energy is for the metal's energy to be minimized. (Low temperature diminishes the effect of the system having a high entropy. Minimizing entropy is not sufficient, because low entropy with high energy could prevent minimization of the Helmholtz free energy.) Hence, at low temperatures, the spins of the metal can become ordered, giving rise to spontaneous magnetization. When the temperature of the metal is higher, however, the best way for the metal to have a minimum Helmholtz free energy is for the metal's entropy to be maximized (because the high temperature enhances the effect a high entropy has upon lowering the system's energy). Maximized entropy favors a random alignment of the spins and can result in a loss of spontaneous magnetization. This is confirmed by experiment: if one heats an iron magnet to a temperature of about 1043 Kelvin, it becomes demagnetized. Similar reasoning applies to the difference between the solid and non-solid phases of a fluid, where the solid phase corresponds to not much random motion of the particles and hence more order in the fluid, and the non-solid phase corresponds to more random motion of the particles and hence more disorder in the fluid.

In addition to this very general, very qualitative understanding of phase transitions, thermodynamics provides a way of representing the occurrence of a phase transition. It is an experimental fact that, at a thermal phase transition – a phase transition due to a change in a system's temperature – it is possible to put heat into a system without increasing its temperature. For instance, if an uncovered pot of water at room temperature is placed on a gas stove, eventually the flame beneath the pot will raise the temperature of the water to 373 Kelvin, and the water will boil; but if the boiling water continues to be heated by the same flame, its temperature will not increase. The amount of heat required to increase the temperature of a system by 1 Kelvin is known as the heat capacity of the system. So, since the temperature of boiling water does not increase at all despite its continued heating, the heat capacity of boiling water is infinite. This is also true for the heat capacity of melting ice. Generally, the heat capacity of a system that is undergoing a thermal phase transition is infinite. (More carefully, the heat capacity of a system that is undergoing a thermal phase transition is undefined.) The same kind of discontinuity in heat capacity occurs in metals when they become spontaneously magnetized, and in alloys such as brass (an alloy of copper and zinc) for their order-disorder transition.

The Gibbs free energy of a system provides a second indicator for when a system undergoes a phase transition. The Gibbs free energy of a system, G, is something like the amount of energy required to make the system out of nothing, less the amount of energy contributed by an environment at temperature T, plus the amount of energy required to make room for the system of volume V in an environment at pressure P:

$$\begin{aligned} G &= E - TS + PV \\ &= F + PV. \end{aligned}$$

Like the Helmholtz free energy, the Gibbs free energy is minimized for a system that is in equilibrium at a fixed pressure. It can be experimentally determined that the Gibbs free energies for different phases of a system can be considerably unequal. For instance, at a pressure of 1 bar and a temperature of 298 Kelvin (room temperature), the Gibbs free energy for a mole of diamond is greater than the Gibbs free energy for a mole of graphite by about 2900 Joules. (See Figure 3.2.[14]) It can also be experimentally determined that the minimum Gibbs free energy of a system is not always the Gibbs free energy associated with one specific phase of the system. For instance, at pressures lower than about 15 kilobars, the Gibbs free energy of graphite at room temperature is lower than the Gibbs free energy for diamond at room temperature; but at pressures greater than about 15 kilobars this is reversed, so that the Gibbs free energy of diamond is lower than that of graphite. Hence, since the Gibbs free energy of a collection of carbon is minimized, there is a sudden change in the Gibbs free energy of carbon at pressures of about 15 kilobars at room temperature. That

[14]This figure is taken from [104], p. 170, Figure 5.15.

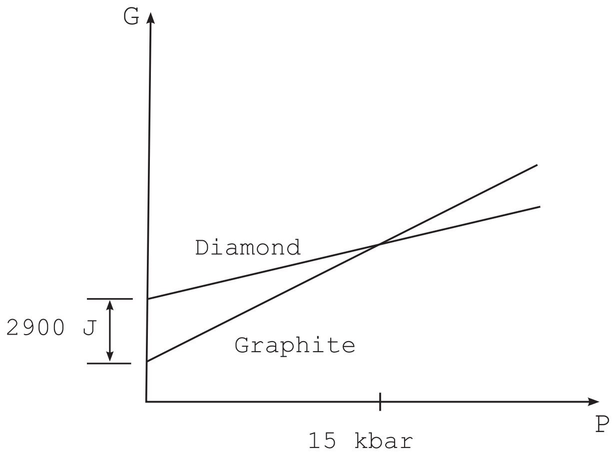

Figure 3.2: Gibbs Free Energy-Pressure Diagram for Diamond and Graphite, at Room Temperature

is, room temperature carbon undergoes a phase transition at a pressure of about 15 kilobars.

This sudden change in the Gibbs free energy of carbon, when the system is undergoing a transition between its diamond and graphite phases, is characteristic of many systems in which phase transitions occur, including water and iron. Such a sudden change in the Gibbs free energy of a system corresponds to a singularity in the Gibbs free energy of the system. (A singularity is a point at which a function is not defined, fails to be differentiable, or fails to yield a unique output for a given set of parameter values.) At a phase transition the first derivative of the Gibbs free energy with respect to the pressure of the system is not well-defined. For instance, the volume of a mole of graphite is about 5.31×10^{-6} cubic meters, whereas the volume of a mole of diamond is about 3.42×10^{-6} cubic meters. Since the volume of a system is the partial derivative of the Gibbs free energy of the system with respect to its pressure, the first derivative of the Gibbs free energy of carbon is not well-defined at the graphite-diamond transition.

Thermodynamics not only provides a rough qualitative understanding of phase transitions, but also shows that phase transitions occur in a system when a thermodynamic property of the system becomes singular. For instance, the heat capacity of water is not well-defined at the liquid-gas phase transition. Nor is the heat capacity of a ferromagnet (such as iron) well-defined when the iron becomes spontaneously magnetized, as a later discussion of the Ising model will show. The Gibbs free energy of carbon is discontinuous at the graphite-diamond phase transition. All of this suggests that phase transitions should be represented as singularities in certain thermodynamic properties. Moreover, this suggests that statistical mechanics should represent phase transitions as singularities, so that the singularities in statistical mechanical functions can correspond to singularities in thermodynamic properties.

3.1.2 Statistical Mechanical Point of View

Statistical mechanics aims to describe the properties and behaviors of very large systems in terms of the microscopic states of these systems. Most, if not all, phase transitions occur in very large systems. Accordingly, the study of phase transitions falls within the purview of statistical mechanics.

Given the thermodynamic approach to phase transitions, it is natural for statistical mechanics to want to identify phase transitions as singularities in a function known as the partition function.[15] The reason for this is that the natural logarithm of the

[15]This approach to the statistical mechanical analysis of phase transitions is known as the Lee-Yang theory of phase transitions; for an overview of this theory, see [10]. For the theory as originally given by Lee and Yang, see [124] and [68]. For an application of this theory to an account of phase transitions in magnets, see [25] and Chapter 10 of [53].

partition function of a system, Q, is proportional to the Helmholtz free energy of the system multiplied by $-\beta$:

$$\ln Q = -F\beta.$$

(β is the system's inverse temperature, by definition equal to $1/k_B T$, where T is the system's temperature and k_B is Boltzmann's constant.) The connection between the logarithm of the partition function and the Helmholtz free energy times $-\beta$ is justified by the fact that both quantities behave in the same way: both decrease in the same way whenever the entropy of a system increases. It is a result from thermodynamics that a system undergoes a phase transition just in case its Helmholtz free energy contains a singularity. This result is connected to the singularities in the heat capacity or Gibbs free energy of a system during a phase transition. It turns out that identifying phase transitions in a system as singularities in the partition function of the system will not work. Before elaborating on this claim, however, it will be helpful to say a bit more about what the partition function is.

The partition function is a measure of the number of microstates accessible to a system in thermal equilibrium at a specific temperature.[16] For instance, the partition function for an isolated system of non-interfering particles with energies E_1 and E_2, respectively, is

$$\sum_s \exp^{-\beta[E_1(s)+E_2(s)]},$$

[16]The partition function is also a normalization "constant" for the Boltzmann probability distribution, which is a distribution that gives the probability of finding a system at a specific temperature in a specific microstate, given that the system is in thermal equilibrium. The partition function is the denominator for this probability distribution. Statistical mechanics introduces probability distributions since the microscopic state of a system cannot be known with certainty. See [20], pp. 9-10 for a discussion of this.

where the sum is over all possible states s for the composite system. If the particles are distinguishable, if each can occupy exactly one of two total positions, and if the particles have no other variable properties, then there are two possible states for the composite system – viz., s_a and s_b – and the partition function for the system is equal to

$$\exp^{-\beta[E_1(s_a)+E_2(s_a)]} + \exp^{-\beta[E_1(s_b)+E_2(s_b)]}.$$

Generally, the partition function for a system is obtained by multiplying the Hamiltonian (total energy) for the system by a factor of $-\beta$ in order to cancel the dimensional units of the Hamiltonian, exponentiating this product for the system, and then summing the resulting power over all the possible microscopic states of the system. (The partition function contains an exponent, because it is a sum over all possible microscopic states and that sum grows exponentially as the number of possible microscopic states increases.)

For example, a "canonical" system is a system with a fixed temperature and a fixed number of particles that exchanges only heat with its external environment. The partition function $Q(\Lambda, T)$ for a canonical system Λ with temperature T is given as

$$Q(\Lambda, T) = \sum_i \exp{(-\frac{E_i}{k_B T})},$$

where the sum is over each state i with energy E_i. Q is a measure of how many different energy states are appreciably populated when a canonical system is in thermal equilibrium at temperature T; it is a measure of the number of states accessible to a canonical system that is in thermal equilibrium at temperature T. There are, of course, other partition functions for other kinds of systems. For example, there are systems that exchange heat as well as particles with their external environment;

these systems are known as “grand canonical” systems, and their partition function is known as the grand canonical partition function. The partition function is significant, because it can be used to calculate all the macroscopic properties of a system.[17]

As noted, it is natural for statistical mechanics to want to identify phase transitions in a system as singularities in the system’s partition function.[18] (The partition function contains a singularity at a point if it cannot be expanded as a Taylor series around that point.) The main reason for this is that the logarithm of the partition function, $\ln Q$, is equal to $-F/k_BT$. Since k_B is a (non-singular) constant and T is fixed at a specific value in order for the partition function to be well-defined, the partition function contains a singularity just in case the Helmholtz free energy F contains a singularity.

The partition function of an N-particle system with $N < \infty$ is a finite sum of the system’s energy configurations. Since each of these terms is analytic, the partition function is analytic – and so it does not contain any singularity. Similarly, it is a mathematical fact that the Helmholtz free energy of any N-particle system with $N < \infty$ is analytic and hence does not contain any singularity.

In order to obtain a function related to the Helmholtz free energy and partition function that contains a singularity, it is necessary to study the thermodynamic limit,

[17]If a_i is the value of a macroscopic property A of a canonical system in the microscopic state i, and if E_i is the energy of this state, then the average value $\langle a \rangle$, interpreted as the measurable value of A when the system is in thermal equilibrium at temperature T, is $\langle a \rangle = \frac{1}{Q} \sum_i a_i \exp{(-E_i \beta)}$.

[18]The representation of a phase transition as a singularity is a useful way to identify the breaking of symmetries, often connected to the occurrence of phase transitions. Axel Gelfert briefly discusses this connection: “The occurrence of a phase transition is often linked to a failure of one of the phases to exhibit a certain symmetry property of the underlying Hamiltonian. Crystals, for example, by their very lattice structure, break the translational symmetry encountered in the continuum description of fluids; ferromagnets, in addition to the spatial symmetry-breaking due to their crystal structure, are not invariant under rotations in spin space, even though the underlying Hamiltonians describing the system may well be” ([37], p. 4).

in which the system's particle number $N \to \infty$, the system's volume $V \to \infty$, and the density of the system retains its finite, actual value.[19] (The thermodynamic limit is usually interpreted as the limit in which the system's volume and number of particles become infinite, or arbitrarily large. This is what, in the next section of the chapter, I will call interpreting the thermodynamic limit as a distorting idealization.) In the thermodynamic limit, the Helmholtz free energy *per particle* of a system contains a singularity if the system undergoes a phase transition; this is connected to singularities in the *specific* heat capacity (heat capacity *per unit mass*) or Gibbs free energy *per particle* of the system. For an N-particle system, its Helmholtz free energy per particle, f, is defined to be its Helmholtz free energy divided by its number of particles: $f \equiv F/N$. Hence, rather than following thermodynamics in identifying phase transitions as singularities in the Helmholtz free energy of a system, statistical mechanics identifies phase transitions as singularities in a system's Helmholtz free energy per particle.

The reason why statistical mechanics identifies phase transitions as singularities in the Helmholtz free energy per particle rather than the Helmholtz free energy itself (as in thermodynamics) is that phase transitions only occur in the thermodynamic limit, and it is a mathematical fact that in the thermodynamic limit the Helmholtz free energy per particle is well-defined (if the limit exists) but the Helmholtz free energy itself is not well-defined. This fact can be explained in the following manner. An intensive property is a property that does not double when the size of a system doubles. (For instance, temperature, pressure, and density are intensive properties. If

[19] The notation '$N \to \infty$' may be read as 'N goes to infinity' or 'N approaches infinity'. The thermodynamic limit also figures in other important results from statistical mechanics, such as the proof of the equivalence of ensembles and the explanation of Bose condensation. See [114] for a discussion of six results in statistical mechanics that involve an appeal to the thermodynamic limit.

one gas at a certain pressure and temperature and with a certain density is added to another gas with the same pressure, temperature, and density, the resultant gas has the same pressure, temperature, and density.) Non-intensive, or extensive, properties are those that do double when the size of a system doubles; these include mass, volume, number of particles, energy, and entropy.

In the thermodynamic limit, a system's particle number $N \to \infty$. If this limit is interpreted to be a distortion, the thermodynamic limit is the limit in which a system's number of particles becomes infinite (or arbitrarily large). Hence, in the thermodynamic limit, the extensive properties of the system also become infinite (or arbitrarily large), which is to say that they are no longer well-defined. For instance, suppose a system is composed of N particles each of nonzero mass m, such that the total mass of the system is the product of the number of particles in the system and the representative mass of one such particle: $M = Nm$. Then in the thermodynamic limit, the system's total mass $M \to \infty$, since $N \to \infty$. However, since the mass per particle is an intensive property of the system, it remains well-defined in the thermodynamic limit (if the limit exists). Likewise, since the Helmholtz free energy of a system is equal to the product of the Helmholtz free energy per particle and the number of particles in the system, the Helmholtz free energy of a system is not well-defined in the thermodynamic limit, even though the Helmholtz free energy per particle is well-defined in this limit. Since phase transitions only occur in the thermodynamic limit, if statistical mechanics is to indicate the occurrence of a phase transition by a singularity in some well-defined function, then that function must be a function of intensive properties only.

A useful way to illustrate the way in which statistical mechanics indicates the occurrence of phase transitions as singularities, to illustrate the fact that the thermodynamic limit is necessary for such singularities, and to illustrate the fact that the functions in which such singularities occur are functions of intensive properties only, is to consider the Ising model.

3.1.3 The Ising Model

The Ising model captures the qualitative behaviors of ferromagnets, such as iron and nickel (metals that can be magnetized). A distinctive property of ferromagnets is that, below a temperature known as the Curie temperature (which is different for different systems), they are in a ferromagnetic phase with a spontaneous magnetization, and that above this temperature their spontaneous magnetization vanishes and they are in a paramagnetic (non-magnetized) phase. Ferromagnets undergo a ferromagnetic phase transition as their temperature approaches the Curie temperature. The Ising model, an idealized model of real ferromagnets, is able to predict and explain this property of ferromagnets: as the temperature of a system described by the Ising model approaches its Curie temperature, the heat capacity of the system diverges logarithmically, indicating the occurrence of a phase transition. Unfortunately, there is no known solution to equations that describe the 3-dimensional Ising model. So, in the interests of exactitude, and despite the fact that most real systems are 3-dimensional, it is best to focus on the Ising model for 2-dimensional systems, since the relevant equations for this model have known exact solutions.

The 2-dimensional Ising model models a ferromagnet as a lattice consisting of a fixed set of regularly-spaced sites. These sites are connected to each other by "bonds", in such a way that each bond connects exactly two sites and every bond is either parallel or orthogonal to every other bond. (See Figure 3.3.) Sites that are

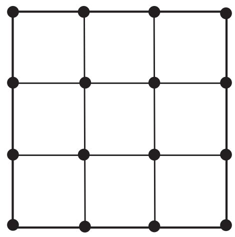

Figure 3.3: 2-Dimensional 4 × 4 Lattice

connected by a single bond are called "nearest neighbor" sites. Each bond is of equal length, so that every site has four nearest neighbors (except for sites at the boundary of the lattice, which are ignored in the thermodynamic limit). Each site has either an "up" spin of +1 or a "down" spin of -1.[20] Only nearest neighbor sites interact in the model. The interaction energy between two sites is either $-J$, if the sites have the same spin, or $+J$, if the sites have different spins; J is a parameter for the strength of the interaction that is positive for ferromagnets.

The 2-dimensional Ising model represents a magnetic material, if one adopts the following correspondences. Each site corresponds to an atom of the material; the spin at a site corresponds to the magnetic moment of each atom; and the restriction to nearest neighbor interactions corresponds to the assumption that inter-atomic interactions are short-range.

[20] If, instead, one supposes that each site is either vacant or occupied, one obtains a lattice gas model, which can be used to model the solid-gas phase transition in fluids.

If the magnetic material is supposed to have $M = N^2$ sites (or, equivalently, N^2 magnetic moments), then the Ising model for the material is an $N \times N$ lattice.[21] There are 2^M ways to configure an $N \times N$ lattice so that each site σ_i is either spin up ($\sigma_i = +1$) or spin down ($\sigma_i = -1$). The total interaction energy of each such configuration is

$$-J \sum_{<i,j>} \sigma_i \sigma_j,$$

where the sum is over nearest neighbor sites. If there is an external magnetic field of strength B present, then the total energy due to the interaction of each site with this field is

$$B \sum_{<i>} \sigma_i,$$

where the sum is over all sites of the lattice; this field tends to align each spin in a direction parallel to the field. Hence, the total energy E for each configuration $\sigma = (\sigma_1, \ldots, \sigma_{N^2})$ of the lattice is the total interaction energy of the configuration less the total energy due to the interaction of each site in the configuration with the external magnetic field:

$$E_\sigma = -J \sum_{<i,j>} \sigma_i \sigma_j - B \sum_{<i>} \sigma_i.$$

When the strength of the external magnetic field is non-zero ($B \neq 0$), the external magnetic field induces magnetization in the material, by aligning most of the magnetic moments parallel to the field and to each other. The strength of this magnetization depends upon the strength of the external magnetic field and the type of material. When the external magnetic field is removed, the material loses its magnetization if its temperature is above the Curie temperature T_c for the material. If, however,

[21] This 'M' should not be confused with the symbol for representing the total mass of a system.

the temperature is below T_c, the material retains a residual magnetization; this is spontaneous magnetization ('spontaneous', because it is present without external influence). When spontaneous magnetization occurs, the material is said to undergo a ferromagnetic phase transition.

The Ising model predicts the existence of this Curie temperature. Each possible configuration σ of an $N \times N$ lattice at a fixed temperature T is assigned a probability $P(\sigma)$, such that

$$P(\sigma) = \frac{\exp^{-\beta E_\sigma}}{Q},$$

where the inverse temperature $\beta \equiv (k_B T)^{-1}$ and Q is the canonical partition function ('canonical', because the lattice size is fixed). Q is an exponentiation of the total energy of the lattice, summed over the 2^M possible configurations of the lattice:

$$Q = \sum_\sigma \exp^{-\beta E_\sigma}.$$

The partition function contains information about the global behavior of the lattice and is, accordingly, the place to look for whether the lattice has a Curie temperature at which the ferromagnetic phase transition occurs.

It is possible to show, for an $N \times N$ lattice with external magnetic field $B = 0$, that[22]

$$\frac{\ln Q}{N^2} = \ln 2 + 2(1 - \frac{1}{N}) \ln\cosh(J\beta)$$
$$+ \frac{1}{8\pi^2} \int_0^{2\pi} \int_0^{2\pi} d\varepsilon d\eta \ln[\cosh^{-4} 2J\beta([\cosh 2J\beta]^2 - [\sinh 2J\beta]^2[\cosh\eta + \cosh\varepsilon])],$$

where $0 \leq \varepsilon \leq 2\pi, 0 \leq \eta \leq 2\pi$. Apart from a factor of $-\beta$, this is the Helmholtz free energy *per site* (or per spin), f, for an $N \times N$ lattice, defined by the bridge law

$$f = -\frac{1}{\beta}\frac{\ln Q}{N^2}.$$

[22]This follows [26].

There are no singularities in f for $N < \infty$: it is continuous over all temperatures (for all values of β). However, in the thermodynamic limit $N \rightarrow \infty$, and only in this limit, there is a singularity in f. This is why the thermodynamic limit is an ineliminable idealization in the explanation of ferromagnetic phase transitions: only in this limit does something that represents a phase transition occur (namely, a singularity in the Helmholtz free energy per site).

Apart from a factor of $-\beta$, the Helmholtz free energy per site in the thermodynamic limit is

$$\lim_{N\rightarrow\infty} \frac{\ln Q}{N^2} = \ln 2 + \frac{1}{2\pi} \int_0^\pi \int_0^\pi d\varepsilon d\eta \ln[(\cosh 2J\beta)^2 - (\sinh 2J\beta[\cosh \eta + \cosh \varepsilon])].$$

Note that, in this limit, the free energy itself, $F = fN^2$, is not well-defined; this is to be expected, since the free energy F is an extensive property of the system. It is notable that, since singularities occur in the free energy per site only in the thermodynamic limit, and since the free energy itself is not well-defined in the thermodynamic limit, the logarithm of the partition function is not well-defined in the thermodynamic limit either, because the logarithm of the partition function is proportional to the free energy divided by temperature.

Defining $k \equiv (\tanh 2J\beta)(2\cosh 2J\beta)^{-1}$, the free energy per site in the thermodynamic limit may be rewritten as

$$\lim_{N\rightarrow\infty} \frac{\ln Q}{N^2} = \ln 2 + \ln(\cosh 2J\beta) + \frac{1}{2\pi} \int_0^\pi \int_0^\pi d\varepsilon d\eta \ln(1 - 2k[\cosh \eta + \cosh \varepsilon]).$$

The logarithm of the third term on the right hand side of this equation may be expanded in powers of k, to yield

$$-f\beta = \ln 2 + \ln(\cosh 2J\beta) - \sum_{n=1}^{\infty} (\frac{(2n)!}{(n!)^2})^2 k^{2n}.$$

The first two terms on the right hand side of this expansion are analytic for all temperatures. The series of the third term, however, diverges at $k = 1/4$ (for $k > 0$, $J > 0$); this corresponds to a temperature

$$T_c = \frac{2J}{k_B \ln(\sqrt{2} + 1)}.$$

This is the Curie temperature for an $N \times N$ Ising model. This corresponds to a logarithmic divergence in the specific heat at the temperature T_c, which indicates the occurrence of a phase transition at temperature $T = T_c$. (Specific heat is also an intensive property, equal to the amount of heat required to raise the temperature of the system, per Kelvin of temperature increase, per unit mass.)

3.2 Criticisms

The explanation that statistical mechanics gives of the occurrence of phase transitions is an idealized explanation. For statistical mechanics explains the occurrence of phase transitions for systems that are in the thermodynamic limit. Hence, if the thermodynamic limit is a distorting idealization, statistical mechanics treats real systems as if they have an infinite number of components.[23] For example, any real glass of pure ice contains only finitely many hydrogen and oxygen atoms; yet statistical mechanics must treat the ice in that glass as if it contains infinitely many hydrogen and oxygen atoms in order to explain the melting of the ice in the glass. If the ice in the glass is taken to have only finitely many atoms, statistical mechanics is unable to

[23]Strictly speaking, this is not correct. For a system to be in the $N \to \infty$ limit is not for it to have infinitely many particles, but for the number of particles to be arbitrarily large. ('Arbitrarily large' is to be understood in terms of the definition of a limit.) The phrase 'the number of particles is infinite' should be understood as shorthand for the technically more correct phrase 'the number of particles is arbitrarily large', the phrase 'infinitely many particles' should be understood as shorthand for 'arbitrarily many particles', etc. So, too, saying that real systems do not contain infinitely many particles should be taken as a quick way of saying that real systems do not have arbitrarily large numbers of particles.

describe the melting of the ice, since the free energy per particle for the system will fail to develop singularities.[24] Since real systems have only finitely many components, statistical mechanics describes idealized versions of real systems when it explains the occurrence of phase transitions in those real systems. Since the content of the statistical mechanical explanation of phase transitions involves an idealized description, it is legitimate to ask how this explanation explains what happens in the real world. No extant account of idealized explanation that takes idealizations to be distortions has an answer to this question.

First, consider Laymon's account. According to Laymon, an explanation that involves an idealized description has two components, an idealized sketch and a modal auxiliary. The statistical mechanical explanation provides an idealized sketch, to the effect that phase transitions in a system occur when the Helmholtz free energy per particle of the system develops a singularity. This sketch is idealized, because it treats real systems as if they contain an infinite number of components. Any improvement upon this idealized description that renders it more realistic would have to consider a system that contains only a finite number of components.[25] Hence, the modal auxiliary for the statistical mechanical explanation must be an argument to the effect that the Helmholtz free energy per particle for a system with finitely many components can develop singularities. Yet treating systems as if they have infinitely many components is necessary for the success of the statistical mechanical

[24] For a discussion of attempts to represent phase transitions as something other than singularities, see Chapter Five.

[25] Any other improvement sidesteps the main issue. Laymon requires the series of improvements to the idealized description to produce a monotonic convergence to a correct description; and a correct description requires that the system of interest contain only finitely many particles, if the system is a real system. If the idealized description contains an idealization that cannot be improved upon, the monotonic convergence to a correct description can only be partial, never complete.

explanation. Hence, there cannot be the requisite modal auxiliary for the idealized statistical mechanical explanation; improving the system, so that it has finitely many particles, entails the failure of this account of phase transitions. Thus, Laymon's exigent account of idealized explanation does not show how the statistical mechanical account of phase transitions is explanatory.

Second, consider Rueger and Sharp's account. According to Rueger and Sharp (RS), an explanation that involves an idealized description shows that real systems qualitatively approximate idealized versions of themselves. To show that a real system qualitatively approximates an idealized version of itself is to show that either the law that describes the idealized system is structurally stable, or the law family to which that law belongs is structurally stable as a family. The applicability of RS's account requires the real system to be some perturbed version of the idealized system. If this requirement is to be satisfied in the statistical mechanical case, the idealized system must be perturbed with respect to the parameter that controls for the number of particles in the system, N. Since only small perturbations are allowed, it seems that no real system, with finitely many particles, is among the systems that are perturbed versions of the idealized system. So RS's analysis seems not to apply to the statistical mechanical account of phase transitions.

Suppose, however, for the sake of argument, that a perturbation from an infinite particle number to a finite particle number counts as a small perturbation. Then RS's account still does not apply. The idealized system that appears in the statistical mechanical account of phase transitions is in the thermodynamic limit. Hence, every perturbation of this system with respect to the number of particles yields a system with finitely many particles.[26] Since singularities develop only in the thermodynamic limit, when the number of particles is infinite, no Helmholtz free energy per particle function for any appropriately perturbed version of the idealized system contains any point of singularity, even though the Helmholtz free energy per particle function for the idealized system contains points of singularity. (In this case, a perturbation of the idealized system is appropriate if it is with respect to the number of particles of the idealized system.) There is no homeomorphism that transforms the phase space portrait for a Helmholtz free energy per particle function that contains a singularity into one that does not. Hence, the Helmholtz free energy per particle function for the idealized system is structurally unstable.

Furthermore, the law family for the Helmholtz free energy per particle function of this idealized system is not structurally stable as a family. The Helmholtz free energy per particle function for the idealized system is a function of N, V, and T. The law family for this function is the set of functions generated by allowing N, V, and T to take on values that are small perturbations from the values they have in the thermodynamic limit. Suppose, for the sake of argument, that one member of this law family is the Helmholtz free energy per particle function for the real system, with finite N and finite V. Then the Helmholtz free energy per particle function for this real system is also structurally unstable, since the Helmholtz free energy per particle function for the perturbation of the real system that takes N to infinity contains

[26] Perturbing the system so that it has uncountably many particles, rather than a countable infinitude of particles, results in a system that is still idealized. No real system has uncountably many particles. Such a perturbation cannot show that the real system has the same features as the idealized system with (countably) infinitely many particles, since the real system will not be one of the perturbed versions of the idealized system.

points of singularity whereas the Helmholtz free energy per particle function for the real system contains no such points.

Yet the Helmholtz free energy per particle function for the real system is not structurally unstable in the same way that the partition function for the idealized system is structurally unstable. The former is structurally unstable because a certain perturbation with respect to N leads to a system with a Helmholtz free energy per particle function that contains singularities not present in the Helmholtz free energy per particle function for the unperturbed system. The latter is structurally unstable for the opposite reason: there is a perturbation with respect to N that results in a system with a Helmholtz free energy per particle function that lacks the singularities present in the Helmholtz free energy per particle function for the unperturbed system. (Indeed, *every* perturbation with respect to N towards a finite value of N gives this result.) Hence, not every member of the law family for the Helmholtz free energy per particle function of the idealized system is structurally unstable in the same way. Since the Helmholtz free energy per particle function for the idealized system is not structurally stable, not every member of the law family for this function is structurally stable. Consequently, the law family for the Helmholtz free energy per particle function of the idealized system that appears in the statistical mechanical account of phase transitions is not structurally stable as a family.[27]

[27]Here there is an appearance of disagreement with Rueger's analysis (see [99], pp. 484-485). Rueger claims that, although the systems in which phase transitions occur are structurally unstable, the family of such systems is structurally stable as a family. Rueger's analysis of phase transitions is a thermodynamical analysis (he calls it "phenomenological"), because he appeals to properties of the van der Waals equation. I provide a different analysis – what might be called a statistical mechanical analysis – insofar as I appeal to properties of the Helmholtz free energy per particle. I claim that this function for a real system is not structurally unstable *in the same way* that the Helmholtz free energy per particle function for the idealized system is structurally unstable. Given the focus of Rueger's analysis, he does not address the fact that phase transitions occur only for systems that exist in the thermodynamic limit. Although I concede that systems that exist in the thermodynamic limit

Since the Helmholtz free energy per particle function of this idealized system is structurally unstable, and since the law family for this function is structurally unstable as a family, the idealized system does not qualitatively approximate any real system with finitely many particles. The statistical mechanical account of phase transitions requires the appeal to idealized systems with infinitely many particles. Hence, since such systems fail to qualitatively approximate real systems, RS's account of idealized explanation does not show how the statistical mechanical account of phase transitions is explanatory.

Finally, consider Kitcher's account. According to Kitcher, an explanation that involves an idealized description has (as I put it earlier) a prologue and an epilogue. The epilogue for the statistical mechanical explanation of phase transitions is supposed to show the ways in which real systems differ from systems that have an infinite number of particles. One of the ways these systems differ, according to statistical mechanics, is that real systems do not undergo phase transitions. Hence, the epilogue for this explanation cannot show how phase transitions in real systems differ from phase transitions in the idealized system, because according to statistical mechanics there are no phase transitions in real systems. There are only phase transitions in idealized systems that exist in the thermodynamic limit. This is not a surprising result, considering that there is no justification, in the sense that Kitcher requires, for the prologue of such an explanation. For the idealization of the number of particles in a system has a high probability (actually, a probability of one) of making a significant difference to the phenomenon of interest, the occurrence of phase transitions.

and exhibit phase transitions are structurally stable as families, I do not concede that the family containing both these idealized systems and real systems is structurally stable as a family.

This shortcoming of Kitcher's account is not due to the particular conditions he requires of the prologue and epilogue. If one were to adapt the innovations of the other accounts of idealized explanation to Kitcher's basic model, the resultant account would still fail to apply. For example, following Laymon, one might eliminate the epilogue and replace Kitcher's justification of the prologue with one that shows there is a modal auxiliary for the idealized description in the prologue; or, following RS, one might replace this justification with one that shows the idealized description qualitatively to approximate the correct description, and replace Kitcher's epilogue with one that identifies the topological properties shared by the real system and its idealized version. These alterations to Kitcher's account would not allow it to accommodate the statistical mechanical account of phase transitions, for the reasons given against the other accounts.

3.3 Irreversibility and the Boltzmann-Grad Limit

When a balloon full of helium pops in the corner of a room, the helium is initially near the corner; over time, it spreads throughout the room. Although it is possible for the helium to reunite in the corner of the room, we never observe this: the dissipation of the helium is irreversible. Cigarette smoke spreading throughout a room is similarly irreversible.

When an insoluble drop of black ink is stirred in a glass of water, the ink distributes itself evenly throughout the water, turning it a uniform gray. Although it is possible to stir the water in a way that results in a reformation of the original ink droplet, we never observe this: the spread of the ink is irreversible. Sugar dissolving in water is similarly irreversible.

When a warm body comes into contact with a cold one (and is otherwise isolated from the larger environment), heat flows from the warm body to the cold one, until the two bodies reach the same uniform temperature. Although it is possible for heat to spontaneously flow back to the originally warm body, we never observe this: the heat flow is irreversible.

These examples, and others like them, demonstrate the pervasiveness of irreversible processes in our experience of the world. Closed (or isloated) systems tend towards states of increasing entropy and, having reached an equilibrium state, stay there, in the sense that a return to their original non-equilibrium state is so extremely improbable as never to be observed by us. This generalization forms the content of the second law of thermodynamics.

The second law provides a phenomenological explanation of why some processes are irreversible – phenomenological, because the second law does not show how the microscopic behavior of systems leads to irreversibility. The aim of statistical mechanics is to provide this further explanation. There are several competing accounts for how this explanation goes (see [108], pp. 246-279). Omitting the details, interventionist accounts find the origin of irreversibility in the interaction between a system and its environment. Gibbsian accounts find the origin of irreversibility in the sequence of coarse-grained states of a system being overwhelmingly likely to approach a coarse-grained equilibrium state. Boltzmannian accounts find the origin of irreversibility in the overwhelming likelihood of a system's time evolution being covered by some version of what is known as the Boltzmann equation.

There are other accounts besides these three, and unfortunately a discussion of them all is beyond the bounds of this project. Here the focus is on Boltzmannian accounts, which seek an explanation of irreversibility in the Boltzmann equation. The main reason for focusing on this approach is the high degree of mathematical rigor with which its proponents have been able to develop the account. The reason for neglecting the interventionist approach is that it would be nice to have an account of irreversibility even for isolated systems. The reasons for neglecting the Gibbsian approach are more complicated. Here is a sketch of one reason for doing so. The Gibbsian and Boltzmannian approaches both account for irreversibility by showing that the entropy of a non-equilibrium system is extremely unlikely to decrease over time. The two approaches work with different notions of entropy, the Gibbs entropy and the Boltzmann entropy, respectively. The Gibbs entropy is a property of a collection of initially identical systems, whereas the Boltzmann entropy is a property of individual systems. Since the explanandum for an account of irreversibility is why the entropy of an individual system never decreases, the Gibbsian account appears to focus on the wrong explanandum. Also, unlike the Boltzmann entropy, the Gibbs entropy does not change over time, and so cannot properly characterize an approach to equilibrium (see [67], p. S349; [92], pp. 126-130).[28]

The discussion of the Boltzmannian account of irreversibility proceeds as follows. I introduce some technical background, including the notions of Γ-space and μ-space, total and partial probability density functions, the Liouville equation and accompanying theorem, and the BBGKY hierarchy. I discuss the Boltzmann equation, including its derivation from the principles of classical mechanics and two further, fundamental assumptions, known as the assumption of molecular chaos and the idealization

[28] For further reasons in favor of the Boltzmannian approach over the Gibbsian approach, see [15]. I do not mean to claim that the Gibbsian approach lacks the resources to address these criticisms; I only claim that the approach seems to succumb to such criticisms.

of the Boltzmann-Grad limit. I show how these assumptions are required to avoid two classical objections to the Boltzmannian approach, known as the reversibility paradox and the recurrence paradox. This clarifies the ineliminability of these further assumptions, and especially the ineliminability of the Boltzmann-Grad limiting idealization. Following all of this, I show that the philosophical accounts of idealized explanation (from the previous chapter) do not show how the Boltzmannian account of irreversibility is explanatory.

3.3.1 Background

Classical statistical mechanics describes the state of a system of N (structureless) particles in terms of $2N$ 3-dimensional vector functions. N generalized center-of-mass coordinates $q_1, q_2, \ldots, q_N \equiv q$ give the positions of each of the N particles; and N conjugate momenta coordinates $p_1, p_2, \ldots, p_3 \equiv p$ give the conjugate momenta of each particle, (Each q_i and p_i is a 3-dimensional vector function).[29] The Hamiltonian H of the system, equal to the total kinetic and potential energy of the system, determines the motion of these particles, according to the equations[30]

$$\left.\begin{aligned} \dot{q}_i &= \partial H / \partial p_i \\ \dot{p}_i &= -\partial H / \partial q_i \end{aligned}\right\} i = 1, 2, \ldots, N.$$

These $2N$ first-order differential equations are known as Hamilton's equations of motion; they determine the temporal variations in the $2N$ functions $p_i(t)$ and $q_i(t)$.

The state of an N-particle system can be represented as a "phase point" in a $6N$-dimensional Γ-space with $6N$ mutually orthogonal axes $q_{x1}, q_{y1}, q_{z1}, \ldots, q_{xN}, q_{yN}, q_{zN}$

[29] $p_i \equiv \frac{\partial L}{\partial \dot{q}_i}$, where L is the Lagrangian of the system, i.e., the total kinetic energy of the system minus its total potential energy.

[30] I assume here and throughout that H does not depend on any time derivatives of q or p, so that these equations are time-reversal invariant.

and $p_{x1}, p_{y1}, p_{z1}, \ldots, p_{xN}, p_{yN}, p_{zN}$. ($\Gamma$-space is a kind of phase space.) As the system changes over time, the phase points representing the state of the system trace out a trajectory in this Γ-space. Hamilton's equations of motion determine the shape of this trajectory. (See Figure 3.4.[31])

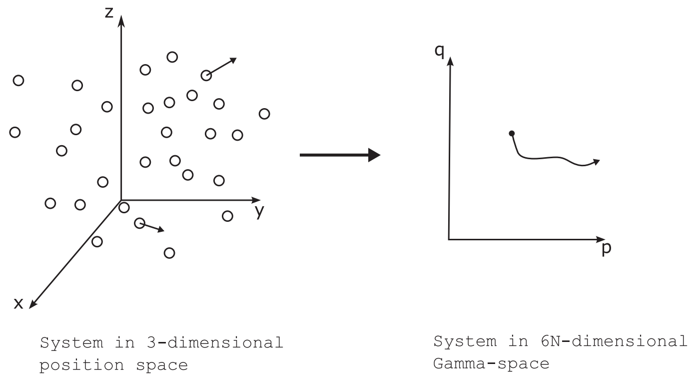

Figure 3.4: Relation between Position-Space and Gamma-Space

Different phase points represent the states of different systems (or different states of the same system). These phase points do not interact with each other; each represents the state of an individual (isolated) system, and the trajectory associated with each phase point represents the time evolution of each system. Generally, the same macroscopic state of a system is compatible with many different N-particle microscopic states of the system, the phase points of which tend to be "near" each

[31]The Boltzmann equation operates in a 6-dimensional phase space known as "μ-space". μ-space has three position dimensions and three momentum dimensions; each particle of a system is represented by a point in μ-space. I would have liked to include a diagram of μ-space, but I'm not sure how to visually represent six dimensions.

other in Γ-space. The collection of microscopic states that correspond to the same macroscopic state is known as a Gibbs ensemble.[32] (See Figure 3.5.)

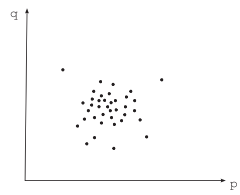

Figure 3.5: Gibbs Ensemble of Phase Points in Gamma-Space

If the Gibbs ensemble is sufficiently large (typically, infinitely many members is sufficiently large), it is possible to describe the distribution of the corresponding phase points in Γ-space by a continuous function of q and p known as a (ensemble) density function. This density function $\rho_N = \rho_N(q, p, t) \prod_{i=1}^{N} dq_i dp_i$ gives the number of phase points in a Gibbs ensemble that, at a time t, occupy the volume between (q, p) and $(q_i + \prod_{i=1}^{N} dq_i, p_i + \prod_{i=1}^{N} dp_i)$ in the Γ-space. (The subscript 'N' on ρ_N is a reminder that each system in the Gibbs ensemble has N particles.) The density function imposes a "coarse-graining" on the Γ-space, since it provides information about volume elements of the Γ-space rather than phase points in the space.[33] The density function changes in time according to Hamilton's equations of motion. It is provable that the density of phase points for a given Gibbs ensemble remains constant

[32] The measure of a (Lebesgue measurable) set of phase points A in Γ-space is defined as $\mu(A) \equiv \int_A dq_1, \ldots, dq_N, dp_1, \ldots, dp_N$.

[33] This is an ad hoc procedure from the point of view of classical mechanics; but there are good quantum mechanical reasons for imposing this coarse-graining, since the exact positions and momenta of every particle in an N-particle system are never simultaneously defined. The complications that arise due to a quantum treatment of an N-particle system are interesting, but they are also irrelevant to present concerns.

over time:

$$\frac{\partial \rho_N}{\partial t} = 0.$$

This is known as Liouville's theorem, and will be important later.

A normalized density function $f_N = f_N(q, p, t)$ can be obtained by dividing the density function ρ_N by the number of members of the Gibbs ensemble. This normalized function is known as the probability density function, since $\int \int f_N \prod_{i=1}^{N} dq_i dp_i = 1$. (Each integral $\int \prod_{i=1}^{N} dx_i$ denotes a $3N$-dimensional integral over each $x_i, i = 1, 2, \ldots, N$.) The probability density function f_N gives the probabilities of finding each particle of an N-particle system within any given region of a 6-dimensional position and momentum space (known as a μ-space), given the exact positions and momenta of the other $N-1$ particles in that space. The function f_N can change in time due to collisions between particles, or due to the motion of particles through space. The equation that governs the time evolution of the probability density function is known as Liouville's equation:

$$\frac{df_N}{dt} = -\sum_{i=1}^{N} [\frac{\partial H}{\partial p_i} \frac{\partial f_N}{\partial q_i} - \frac{\partial H}{\partial q_i} \frac{\partial f_N}{\partial p_i}].$$

According to Liouville's theorem, the probability density function is constant over time: $df_N/dt = 0$. Note that Liouville's equation is time-reversal invariant: if $f_N(q(t), p(t), t)$ is a solution to Liouville's equation, so is $f_N(q(-t), -p(-t), -t)$. This is also important later.

Liouville's equation entails a hierarchy of equations known as the BBGKY hierarchy (named after Bogoliubov, Born, Green, Kirkwood, and Yvon, who independently derived it). Obtaining this hierarchy requires the introduction of partial probability density functions. The function f_N gives the probability of finding each individual particle in some region of μ-space, given the positions and momenta of every other particle in that space; this is the same as giving the probability of finding a system with N particles in some region of Γ-space. A partial probability density function f_s, in contrast, gives the probability of finding $s < N$ randomly chosen particles of an N-particle system in the subregion $\prod_{i=1}^{s} dq_i dp_i$ of Γ-space. This partial function is obtained by integrating over the region of Γ-space associated with the other $N-s$ particles of an N-particle system. Accordingly, the partial probability density function f_s is defined as

$$f_s(x_1, x_2, \ldots, x_s, t) \equiv \int f_N(x_1, \ldots, x_s, \ldots, x_N, t) \prod_{i=s+1}^{N} dx_i,$$

where x_i specifies, in μ-space, the generalized coordinates and conjugate momentum of one particle in an N-particle system. So, for example, $f_1(x_1, t)dx_1$ gives the probability of finding one randomly chosen particle in the volume of μ-space between x_1 and $x_1 + dx_1$ at time t. Note that f_s for $s = N$ just is the original, non-partial probability density f_N for an N-particle system.

For $s < N$, the partial probability density function f_s gives less information about an N-particle system than does f_N. This can be seen with a toy example due to Michael Peters ([92], pp. 60-61). Consider a sprocket that has some radius r and mass m. (See Figure 3.6.) The phase space for this sprocket system has two dimensions, one axis for the radius of a sprocket and an orthogonal axis for its mass. The probability density for finding the sprocket with a radius between r and $r + dr$, and a mass between m and $m + dm$, respectively, is $P(r, m)drdm$. A partial probability distribution function for finding the sprocket only with a certain radius is, accordingly, $P(r) = \int P(r, m)dm$, where the integral is over all of the m-space. $P(r)$ does not contain information about the mass of the sprocket. Whether $P(r)$ is suitable for use in any given situation depends upon the relevance of the sprocket's mass in that situation, and this, in turn, depends upon the "physics" of the situation.

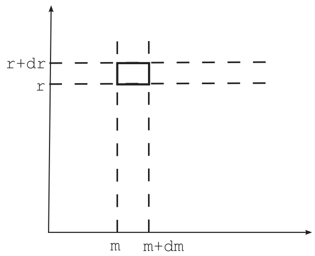

Figure 3.6: Phase Space for a Sprocket

Under certain assumptions (about interaction potentials, etc), the Liouville equation entails a set of N integro-differential equations. The first of these describes the time evolution of the 1-particle partial probability density function; the second, the time evolution of the 2-particle partial probability density function; and so on, up to the N-particle probability density function, which happens to be the Liouville equation. This set of equations is known as the BBGKY hierarchy. Notably, the equations in this set for $s < N$ are not closed: the 1-particle partial probability density function depends upon the 2-particle partial probability density function, which in turn depends upon the 3-particle partial probability density function, and so on. The 1-particle partial probability density function of the BBGKY hierarchy depends upon other partial probability distribution functions, because the probability of finding one particle in some subregion of Γ-space is affected by the collisions of that particle with other particles; these collisions create correlations among the particles, and these correlations affect the probability of finding the particle in various subregions of Γ-space.

Whereas the Liouville equation follows from the laws of classical mechanics, the BBGKY hierarchy follows from those laws plus modeling assumptions about the system. These assumptions might involve, for example, treating each particle as a hard sphere that interacts with other particles via a singular pair potential. (Although these modeling assumptions are idealizations, they are not important to the point I want to make with this example. I introduce a more important idealization shortly.)

3.3.2 The Boltzmann Equation

The Boltzmann equation describes how the statistical distribution of particles in a system changes over time. Specifically, the Boltzmann equation describes the time evolution of a 1-particle partial probability density function – it describes the time evolution for the probability of finding one randomly chosen particle of an N-particle system in some subregion of Γ-space.[34] Unlike the time evolution equation for the 1-particle partial probability density function of the BBGKY hierarchy, the Boltzmann equation is a closed equation: the 1-particle probability density function that appears in the Boltzmann equation does not depend upon every other partial probability density function for $1 < s \leq N$. It is possible rigorously to derive the Boltzmann equation

[34] The Boltzmann equation that describes the time evolution of a 1-particle partial probability density function is the ensemble version of Boltzmann's original equation, which gave the distribution function in μ-space for one particle in a gas. The Boltzmann equation has several applications, in addition to providing an explanation of the irreversible approach to equilibrium. Chief among these is its role in obtaining the Navier-Stokes equations of fluid dynamics (among the most important non-linear differential equations in physics) from statistical mechanics, via a method known as the Chapman-Enskog expansion. (For examples of this derivation, see [31], [75], [102].) Generally, the importance of the Boltzmann equation is its capacity to explain the macroscopic behavior of gases in terms of their microscopic components, thereby bridging "the gap between the atomic structure of matter and its continuum-like behaviour at a macroscopic level" ([20], pp. 40-41). The reason it is desirable to derive the Boltzmann equation, rather than take it as an explanatory starting point, is that a derivation of the equation from the properties and behaviors of the microscopic components of (ensembles of) gases provides an explanation of the behaviors of those gases in terms of statistical mechanics. Simply put, the derivation makes explanations that appeal to the Boltzmann equation "deeper" than they would be otherwise, because it grounds those explanations in statistical mechanics, a more fundamental theory.

from the BBGKY equation governing the 1-particle partial probability density function. The derivation proceeds in two main steps.[35]

The first step is to apply the Boltzmann-Grad limit to the equations of the BBGKY hierarchy. (For details, see [21], pp. 164-167.) The Boltzmann-Grad limit is the limit in which $N \to \infty$ and each particle's diameter $\sigma \to 0$ in such a way that $N\sigma^2$ remains finite and non-zero.[36] Speaking of the Boltzmann-Grad limit as a distorting idealization, Harold Grad notes that, in this limit, "the relative size of molecule to scale of the system becomes vanishingly small" ([60], p. 126). That is, the particles' proper volume $N\sigma^3 \to 0$; this means that the particles of a system in this limit occupy a negligible portion of their container's volume. (Samples of real gases for which this is literally true probably exist only in extreme conditions, such as high-altitude aircraft or high-vacuum environments.) In the Boltzmann-Grad limit, it is provable that the set of phase points in Γ-space for which multiple collisions (triple or higher order) occur has (Lebesgue) measure zero (see [21], pp. 14-16). This shows that it is overwhelmingly likely that only binary collisions (between two particles) occur for systems in the Boltzmann-Grad limit.

[35]For the details of this derivation, see [21], especially Chapter IV; the derivations are provided for a system of N identical particles, each of unit mass, that move in a bounded domain of ν-dimensional Euclidean space and interact via a singular pair potential as hard spheres; a more reader-friendly version of this derivation may be found in [22]. For a closely related derivation, see [60] or [40]. Most of these derivations are valid only for short times (e.g., for one-fifth of the mean free time, the average time between two subsequent collisions of a molecule). But there is a derivation that is valid for all times, provided that the system is a gas expanding into a vacuum; see [22]. Finding derivations that are valid for all times under other (more realistic) conditions is currently an open research problem.

[36]Technically, the $N \to \infty$ limit must be taken before, or faster than, the $\sigma \to 0$ limit, since the limits do not commute. If this order is not observed, $N\sigma^2$ goes to zero. It is important that $N\sigma^2$ remain finite and non-zero, because the mean free path is proportional to $V/\pi N\sigma^2$, where V is the volume of the system.

The result of applying the Boltzmann-Grad limit to the equations of the BBGKY hierarchy is a set of equations known as the Boltzmann hierarchy. Like the equations of the BBGKY hierarchy, the equations of the Boltzmann hierarchy are not closed. This is because, in the Boltzmann-Grad limit, the probability of finding a particle in some subregion of Γ-space still depends upon binary collisions that the particle undergoes. A closed equation can be obtained by invoking an assumption of initial molecular chaos; this is the second main step in the derivation of the Boltzmann equation. According to the assumption of initial molecular chaos, the motion of every particle in the system at an *initial* time $t = 0$ is statistically independent of the motion of every other particle in the system at that time:

$$f_s(q_1, p_1, q_2, p_2, \ldots, q_s, p_s, 0) = \prod_{i=1}^{s} f_1(q_i, p_i, 0).$$ [37]

The reason that the right-hand side of this equation is a product among only 1-particle partial probability density functions is that, if the motions of the particles in a system are statistically independent of each other, the location of one particle in the system is uncorrelated with the location of the system's other particles.[38] When set equal to zero, this 1-particle partial probability density function yields the Boltzmann

[37]Note that this assumption is *weaker* than the assumption originally made by Boltzmann, to the effect that molecular chaos is true at all times. Cercignani, et. al. are able to *derive* Boltzmann's original assumption from the assumption of *initial* molecular chaos. This makes it clear that the assumption is consistent with the dynamics of the system: the assumption is not about the way in which the system behaves, but rather about the initial state of the system. Lawrence Sklar, in a different context, concurs with this assessment ([108], p. 224).

[38]For instance, in a two-particle system, the molecular chaos assumption holds that the probability density of finding one particle with a given position and momentum and the other particle with another position and momentum is equal to the probability density of finding the first particle with its position and momentum and the second particle with any position and momentum *times* the probability density of finding the second particle with its position and momentum and the first particle with any position and momentum. Simultaneously throwing a pair of dice is a common example of a system in which something like the initial molecular chaos assumption holds (and continues to hold so long as the dice do not bump into each other or indirectly affect each other through their interactions with the table on which they are thrown).

equation. Cercignani, et. al. provide two justifications for supposing the initial chaos assumption to be true of a system. One is that, in preparing a state, it is practically impossible for us to manipulate each particle of the system; we can only manipulate the system as a whole. Another is that, if the initial state of a system is selected at random, it is overwhelmingly likely (on mathematical grounds) that the initial chaos assumption holds of the system. Cercignani's justifications are epistemic, concerned with the state of our knowledge about interparticle correlations in a system. It seems to me, however, that if a justification is to be given, it needs to be in terms of ontological considerations, about whether there are any interparticle correlations at $t = 0$. (Others justify the assumption *a posteriori*, on the grounds that it leads to an equation that "works"; still others treat it an an axiom, or as a candidate for a brute fact.)

It is also provable that, in the Boltzmann-Grad limit (and under some further assumptions about the existence and smoothness of the limit, and the existence and uniqueness of solutions to the Boltzmann equation and the equations of the BBGKY hierarchy), initial molecular chaos propagates: if the molecular chaos assumption is true at an initial time $t = 0$, then it is true for all subsequent times $t > 0$. This is because, for systems that exist in the Boltzmann-Grad limit, it is overwhelmingly unlikely for any two particles to collide with each other more than once, or for any two particles to collide with each other if they have previously collided with a common particle. So, for example, given that initial molecular chaos propagates, then if a particle A collides with another particle B, it is extremely unlikely for A to collide with a different particle C, lest there be a statistical correlation between B and C, two particles that have not yet collided.

Given the assurance that initial molecular chaos propagates in the Boltzmann-Grad limit, it is possible to obtain the Boltzmann equation from the Boltzmann hierarchy. This is accomplished by substituting $f_1(q_1, p_1)f_1(q_2, p_2)$ for $f_2(q_1, p_1, q_2, p_2)$ in the equation for the $s = 1$-particle partial probability density function in the Boltzmann hierarchy. The result is a 1-particle partial probability density function that is closed, since it does not depend upon any other partial probability density function for $1 < s \leq N$. Consequently, if the assumptions of initial molecular chaos and the Boltzmann-Grad limit, along with other less fundamental assumptions (about inter-particle potentials, particle shapes, etc) hold of a (macroscopically specified) system, and if the basic equations of classical mechanics are correct,[39] then the Boltzmann equation characterizes the time evolution of "almost all" N-particle systems in an initial Gibbs ensemble for that system. ('Almost all' means 'all but a set of (Lebesgue) measure zero'. The systems not so characterized are those in which non-binary collisions or multiple collisions by the same particle (or both) are important.)

3.3.3 The H-Theorem

Systems governed by the Boltzmann equation can be shown to exhibit irreversible behavior. This is the famous result known as Boltzmann's H-theorem.[40] Converting $f_1(q_1, p_1, t)$ to a function of generalized coordinates and velocities $f_1(q_1, v_1, t)$, define a time-dependent function, to be called an H-function, as:

$$H(t) \equiv \int f_1(q_1, v_1, t) \log f_1(q_1, v_1, t) dv,$$

where the integral is over all velocities.

[39] Assuming they are, when quantum effects are negligible.

[40] The presentation of this theorem follows [20], pp. 137-142.

It is provable that if f_1 is a solution to Boltzmann's equation and no heat flows into the walls of the system's container due to particle collisions with the walls, then $dH/dt \leq 0$. That is, under the general conditions stated, the value of $H(t)$ never increases in time. (This is part of the H-theorem.) It is also provable that $H(t)$ decreases monotonically unless the system is in a state of equilibrium, in which case the H-function is constant through time: $dH/dt = 0$. (This is the remaining part of the H-theorem.) The non-increasing nature of the H-function shows that the Boltzmann equation describes an irreversible time-evolution for some (isolated) systems. Indeed, up to a factor of k_B (Boltzmann's constant), differences in the value of H at different times are the same as differences in the thermodynamic entropy as it appears in the second law of thermodynamics. For this reason, the thermodynamic entropy is known as the Boltzmann entropy, and the H-theorem is considered to be a microscopic version of the second law of thermodynamics.[41]

The rigorous derivation of the Boltzmann equation and the H-theorem provide a foundational justification of the second law of thermodynamics.[42] However, whereas the second law of thermodynamics holds that the entropy of an isolated system never decreases, its microscopic version for systems in the Boltzmann-Grad limit holds that it is overwhelmingly likely – but not certain – that the entropy of an isolated system never decreases. The qualification in the microscopic version of the second law is due to the result (from the end of the section "The Boltzmann Equation") that there

[41] I have a reservation about whether it is proper to say that the H-function corresponds to the entropy. More likely, it corresponds to the entropy per particle. Entropy is an extensive property; hence, when $N \to \infty$ the entropy should be undefined. But since the entropy per particle is an intensive property, it can remain well-defined in the $N \to \infty$ limit.

[42] The justification is foundational in the sense that it proceeds from the basic postulates of classical mechanics (plus a statistical assumption of initial molecular chaos); the derivation is rigorous in the sense that it is a mathematically valid derivation that does not rely upon qualitative reasoning.

is a "small" probability, of Lebesgue measure zero, that an isolated system in the Boltzmann-Grad limit, which satisfies the assumption of initial molecular chaos (and various other modeling assumptions), does not obey the Boltzmann equation and thereby does not obey the H-theorem.

The rigorous derivation of the Boltzmann equation and the H-theorem not only provides a foundational justification for the second law of thermodynamics, but also explains why some systems irreversibly approach equilibrium. In agreement with a nomothetic conception of explanation, the Boltzmann equation is law-like, since it follows from the law-like equations of classical mechanics (when conjoined with appropriate additional assumptions). In agreement with a causal conception of explanation, the Boltzmann equation describes the effects of collisions on the time-evolution of a system; since momentum is conserved in these collisions, it is plausible to suppose that the Boltzmann equation describes causal processes and causal interactions; if so, then the H-theorem shows that these causal processes and interactions almost always produce irreversible behavior. In agreement with a unification conception of explanation, the argument patterns used to derive the Boltzmann equation are plausibly taken to be among those that best unify the belief corpus of science in the limit of its rational development, especially given the connection between the Boltzmann equation and fundamental hydrodynamical equations such as the Navier-Stokes equation.

3.3.4 Paradoxes and Ineliminability

The irreversibility latent in the Boltzmann equation has the ring of paradox to it. Although the Boltzmann equation is obtained from Liouville's equation – which (as noted above) happens to be time-reversal invariant – the Boltzmann equation itself is not time-reversal invariant. This feature of the Boltzmann equation prompted two classic objections, known as the reversibility paradox and the recurrence paradox.

The reversibility paradox, due to Loschmidt, is the simplest of the two paradoxes. The Liouville equation, and all other equations of classical mechanics, are time-reversal invariant. It is not possible to derive an equation that is not time-reversal invariant from equations that are all time-reversal invariant. Hence, according to this paradox, the Boltzmann equation is inconsistent with classical mechanics.

The straightforward response to this paradox is to note that it shows the ineliminability of a non-dynamical postulate, asymmetric under time reversal, from the derivation of the Boltzmann equation, if the Boltzmann equation is to be consistent with classical mechanics. Such a postulate appears in the rigorous derivation of the Boltzmann equation: it is the assumption of initial molecular chaos. (The postulate in non-dynamical, because it does not concern the way in which systems develop over time.) The reversibility paradox shows that this assumption cannot be justified on the basis of dynamical considerations alone. (The justification of the assumption of initial molecular chaos is an open foundational problem, which I pass over.) Indeed, the reversibility paradox shows that some sort of statistical assumption is required in order to derive the Boltzmann equation; but it does not show that this assumption is inconsistent with the underlying dynamics of systems governed by the Boltzmann equation.[43] (Incidentally, the reversibility paradox is one reason for being dissatisfied with the Gibbsian account of irreversibility. As Krylov has shown, it is possible to construct a version of this paradox that casts doubt on the tenability of the Gibbsian

[43]Cercignani, et. al. argue that the assumption is consistent with the underlying dynamics in the Boltzmann-Grad limit; see [21], p. 157.

approach; the paradox succeeds even granting the typical statistical assumptions in the Gibbsian account.)

The recurrence paradox, due to Zermelo, is more complicated than the reversibility paradox, because it involves what is known as Poincaré's theorem.[44] This theorem states, roughly, that almost any (isolated) mechanical system with a given fixed and finite energy and finite spatial extension will, after a finite time, return to a state that is arbitrarily close to its initial state. More precisely: For any arbitrarily small compact set A of (Lebesgue) measurable phase points in Γ-space, let B be the subset of A consisting of those points on trajectories that never return to A having once left A; let $\mu(B)$ be the (Lebesgue) measure of the set B. Then

Poincaré's Theorem: $\mu(B) = 0$.

That is, except for a set of phase points of measure zero, all phase points initially in the region A are on trajectories that will return to A after a finite time.[45] This theorem is restricted to those systems with finite spatial extension and finite energy; this ensures the finiteness of the measure of the energy surface for each system and the compactness of the set A.[46] The proof of Poincaré's theorem involves an appeal to Liouville's theorem (noted above), and an appeal to the fact that trajectories of

[44] The discussion of Poincaré's theorem follows [121], pp. 16-20.

[45] Very roughly, Poincaré's theorem is similar to saying that it is nearly impossible to wind an infinitely long string of spaghetti through a finite volume.

[46] A set is compact just in case it is both closed and bounded. Roughly, a set is closed if it is possible for any point in its complement to be changed by a small amount in any direction yet remain in the complement set. And a set is bounded if it has a finite size. An energy surface is that surface in Γ-space to which the motion of a phase point for a system with energy E is confined. This notion may be elaborated in the following way. Energy is a function of the $6N$ variables of the Hamiltonian of an N-particle system. An energy surface is that surface, in an $6N+1$-dimensional space spanned by E and these $6N$ variables of the Hamiltonian, that is described by $E = H(q,p)$. For example, the energy of a one particle system is a function of the 3 position coordinates and three conjugate momentum coordinates of the particle: $E = H(q_x, q_y, q_z, p_x, p_y, p_z)$. So the energy surface for this one-particle system is a surface described by E in a 7-dimensional space spanned by $E, q_z, q_y, q_z, p_x, p_y, p_z$.

different systems in Γ-space cannot intersect each other.[47] It is instructive to examine a (rough) derivation of the proof, in order to understand why the Boltzmann equation is immune to it.

Consider a system with a fixed and finite energy as well as a finite spatial extension and particle number. There are only finitely many ways to arrange a finite number of particles within a finite volume so that the energy of the system retains its given finite value. And since the energy of the system is finite, there are only finitely many different specifications of the momentum of each particle in the system. Since each microscopic configuration of the system corresponds to a unique arrangement of the system's particles and a unique specification of each particle's momentum, the number of possible microscopic configurations of the system is finite. It follows that the number of possible macroscopic states for the system is finite, since each such state is realized by a subset of the possible microscopic system configurations. Moreover, the system's available volume in Γ-space – the volume it is possible for the system to occupy, the system's energy surface – is also finite: the points corresponding to microscopic configurations that represent a given macroscopic state of the system occupy only a finite volume of Γ-space, and the total available volume for the system in Γ-space is the union of the volumes that correspond to each possible macroscopic system state.

Having established the finiteness of the energy surface for a system with a fixed and finite energy and finite spatial volume and particle number, consider a set A of phase

[47]Liouville's theorem is a consequence of Hamilton's equations of motions. Trajectories cannot intersect each other in Γ-space because systems governed by Hamilton's equations of motion are deterministic. Determinism entails that the trajectory of each phase point through Γ-space is unique; but if the trajectories of two phase points were to overlap, their trajectories would not be unique owing to their being a "branching" of the trajectory of each particle at the point of intersection.

points in some region of Γ-space, such that each point in A represents a specific (but arbitrarily selected) possible macroscopic state of the system under consideration. By definition, each point in A represents a system that has the same energy and spatial extension as the other systems represented by points in A (energy and spatial extension are macroscopic properties of the system), even though each such system has a different microscopic configuration. Since each system represented by a point in A has the same finite energy and volume, the set A is compact. In particular, the set A occupies a (proper) subset of the system's available volume in Γ-space (because it represents only one of the possible macroscopic states of the system).

Let B_0 be the non-empty subset of A consisting of all those phase points on trajectories that never return to A having once left A. Assume, for reductio, that the Lebesgue measure of B_0, $\mu(B_0)$, is both finite and non-zero. Let B_1 be the set of points in Γ-space on trajectories that evolve from the phase points in B_0 after a time t. (The definition of B_0 entals that B_0 and B_1 do not overlap.) Let B_2 be the set of points in Γ-space on trajectories that evolve from the points in B_0 after a time $2t$; and in general, let B_i be the set of points in Γ-space on trajectories that evolve from B_0 after a time it. Then every region B_i is non-overlapping with every other region B_j, including B_0.[48] According to Liouville's theorem, if X is a set of phase points in some region of Γ-space at an initial time and X_t is a (different) set of phase points in Γ-space obtained by evolving the points in X for a time t according to Hamilton's equations of

[48]Here is a rough proof. Suppose, for reductio, that (for $i, j > 1$ and $i < j$) B_i and B_j overlap – i.e., that B_i and B_j have at least one point in common. Let the shared point be x. Since $i < j$, and since trajectories in Γ-space cannot overlap (owing to the determinism of Hamilton's equations of motion), all of the points in B_j at time jt are evolved from points that were in B_i at time $it < jt$. Likewise, since $i, j > 1$, all of the points in B_j and B_i are evolved from points that were in B_1 at time t. In particular, x is on a trajectory that was in B_1 at time t. But since $it \neq jt$ and phase space trajectories cannot intersect, x must also be on a trajectory that was in B at the initial time. This contradicts the assumption that t is large enough that B_0 and B_1 do not overlap. QED.

motion, then $\mu(X) = \mu(X_t)$. Hence, $\mu(B_0) = \mu(B_1) = \mu(B_2) = \cdots$. Therefore, since $\mu(B_0)$ is finite and non-zero (by assumption) and the sequence of regions $B_1, B_2, \ldots$ do not overlap, the total measure of the union of these disjoint regions is infinite. This entails that the whole energy surface for the system is unbounded – i.e., that the system's available volume in Γ-space is infinite. But this is impossible, because the measure of A is finite and the system's energy surface is only finite. Thus, $\mu(B)$ must be zero. QED.[49]

The recurrence paradox builds upon Poincaré's theorem. According to the Boltzmann equation, a system (of the appropriate sort) not in equilibrium that has an initial H-function value $H_i(0)$ is overwhelmingly likely to have an H-function value $H_f(t) < H_i(0)$, for every time $t > 0$ – in fact, $H_f(t)$ should decrease monotonically in time until the system reaches equilibrium. Yet, according to Poincaré's theorem, it is overwhelmingly likely that, at some future time t_r (where $t < t_r < +\infty$), the system has an H-function value H_r that exceeds a previous H-function value of the system. Hence, Boltzmann's equation is inconsistent with classical mechanics.

[49]Here is an alternative method of proof. Let A be an arbitrary volume element in Γ-space, with volume V. After a time t, the points in this volume element evolve into another volume element A_t of volume V_t, according to Hamilton's equations of motion. Let Γ_0 be the subspace that is the union of every region A_t for times $0 \leq t \leq \infty$, and let the volume of Γ_0 be Ω_0. Likewise, let Γ_t be the subspace that is the union of every region A_t for times $\tau \leq t \leq \infty$, and let the volume of Γ_t be Ω_t; the points in Γ_t are the points from Γ_0 that have evolved for a time τ. The volumes Ω_0 and Ω_t are finite, because the finite energy and spatial extension of the system confine a representative phase point for the system to a finite region of Γ-space. All of this entails that the region Γ_0 contains Γ_t. After a time τ, Γ_0 will become Γ_t. Hence, by Liouville's theorem, $\Omega_0 = \Omega_t$. Since these volumes are the same, Γ_0 and Γ_t must contain the same set of points (except for a set of measure zero). Specifically, Γ_t contains all of the points in the volume element A (except for a set of measure zero). But since the points in Γ_t are, by definition, the future destinations of the points in A, all of the points initially in A must return to A after a sufficiently long time (except for a measure zero set of points). Since the volume element A can be made arbitrarily small, every point must return arbitrarily close to its initial state after a sufficient amount of time (except for a measure zero set of points).

The rigorous derivation of the Boltzmann equation shows this inconsistency to be merely apparent. In the Boltzmann-Grad limit, the set of points A is not compact.[50] As $N \to \infty$, a system's energy $E \to \infty$, because energy is an extensive quantity. But a system's average energy per particle, E/N, remains constant ($dE/dN = 0$) as $N \to \infty$; it is an intensive quantity. That $E \to \infty$ and $dE/dN = 0$ in the $N \to \infty$ limit is consistent with an arbitrary particle i in the system having its energy $E_i \to \infty$ as $N \to \infty$. And if a particle's energy $E_i \to \infty$ as $N \to \infty$, its momentum $p_i \to \infty$ as $N \to \infty$, in which case at least one coordinate in Γ-space, in the set A, diverges as $N \to \infty$. This possibility entails that the set A is not compact in the $N \to \infty$ limit.[51] And the non-compactness of A allows the system's energy surface to be unbounded. Hence, Poincaré's theorem does not hold in the Boltzmann-Grad limit; the limit secures the consistency of the Boltzmann equation with classical mechanics, at least so far as the recurrence paradox goes.

This is the reason why the Boltzmann-Grad limit is ineliminable to any rigorous derivation of the Boltzmann equation. Appealing to the limit is the only way, consistent with the Boltzmannian account of irreversibility, to exempt the Boltzmann equation from the strictures of Poincaré's theorem. Exemption cannot be claimed on the grounds that non-equilibrium systems reach equilibrium after an infinite amount of time, since the Boltzmann equation describes how non-equilibrium systems approach equilibrium in a finite amount of time. Nor can exemption be claimed on the grounds that systems are never isolated, since the Boltzmann equation applies to isolated systems. Nor can exemption be claimed on the grounds that the duration

[50] See [21], pp. 171-172; [22], p. 56: "the set [A] is no longer compact when $N \to \infty$ and the recurrence time is expected to go to infinity with N (at a much faster rate)". Cercignani does not say why this is true.

[51] I am grateful to Saul Cohen for assistance with this explanation.

it would take a system to recur to a state similar to its initial state is unfathomably lengthy (*contra* Boltzmann's initial reaction to this paradox), since the Boltzmann equation does not predict this recurrence.[52]

3.4 Criticisms

The Boltzmann-Grad limit is an ineliminable component in the Boltzmannian explanation of irreversibility. For this reason, among others, the Boltzmannian account is an idealized explanation. Interpreted as a distorting idealization, the Boltzmann-Grad limit treats real gases as if they contain infinitely many particles. Appeal to the limit replaces an equation governing the time evolution of a partial probability density function that applies to real gases with an equation (the Boltzmann equation) that holds of idealized gases; and this latter equation holds only of idealized gases, because the appeal to the Boltzmann-Grad limit is ineliminable. Since the Boltzmannian account of irreversibility contains an idealized description (viz., the Boltzmann equation), it is legitimate to ask how this account explains what happens in the real world. No extant philosophical account of idealized explanation that takes idealizations to be distortions has an answer to this question. (The criticisms to follow closely resemble those given for the case of phase transitions. For this reason, the following discussion is brisker, omitting tedious details that can be found in the previous criticisms.)

[52]The Kac ring is an enlightening and surprisingly simple model of how time-reversal invariant equations of motion can lead to irreversible behavior, when conjoined with a non-dynamical statistical assumption and a limit similar to the $N \to \infty$ limit. For discussions of the Kac model in the context of the reversibility and recurrence paradoxes, see [121], pp. 23-27; [29], pp. 34-39. While these discussions emphasize the ineliminability of a statistical assumption to the demonstration of irreversible behavior in the Kac model, they do not emphasize the ineliminability of a limiting idealization similar to the $N \to \infty$ limit, even though the Kac model exhibits irreversible behavior only in such a limit.

First, consider Laymon's account. According to Laymon, an explanation that involves an idealized description must have two components, an idealized sketch and a modal auxiliary. The rigorous derivation of the Boltzmann equation and the H-theorem provide an idealized sketch, to the effect that some non-equilibrium systems irreversibly approach equilibrium. The sketch is idealized, because it treats systems as if they exist in the Boltzmann-Grad limit. Any improvement upon this idealized description that renders it more realistic would have to consider a system that contains only a finite number of particles. (Any other improvements sidestep the main issue.) Hence, the modal auxiliary for the Boltzmannian account of irreversibility must be an argument to the effect that a system that does not exist in the Boltzmann-Grad limit can exhibit irreversible behavior. Yet treating systems as if they exist in the Boltzmann-Grad limit is necessary for the success of the Boltzmannian account. Hence, there cannot be the requisite modal auxiliary for the idealized Boltzmannian explanation of irreversibility; improving the system so that it has finitely many particles exposes the Boltzmannian account to the recurrence paradox. Thus, Laymon's exigent account of idealized explanation does not show how the Boltzmannian account of irreversibility is explanatory despite being idealized.

Second, consider RS's account. According to RS, an explanation that involves an idealized description of a real system must show that the real system qualitatively approximates the idealized version of itself. To show this is to show either that the law describing the idealized system is structurally stable, or that the law family to which that law belongs is structurally stable as a family. The applicability of RS's account requires the real system to be some perturbed version of the idealized system. If this requirement is to be satisfied for the Boltzmannian account, the idealized system must be perturbed with respect to the parameter that controls the number of particles in the system, N. Since only small perturbations are allowed, it seems that no real system is among the systems that are perturbed versions of the idealized system. So RS's analysis seems not to apply to the Boltzmannian account of irreversibility.

Suppose, however, for the sake of argument, that a perturbation from an infinite particle number to a finite particle number counts as a small perturbation. Then RS's account still does not apply. The idealized system that appears in the Boltzmannian account of irreversibility exists in the Boltzmann-Grad limit. Hence, every perturbation of this system with respect to N yields a system with finitely many particles, a system that does not exist in the Boltzmann-Grad limit. Since the Boltzmannian account avoids the recurrence paradox only by invoking the Boltzmann-Grad limit, systems that do not exist in the Boltzmann-Grad limit do not exhibit irreversibility. There is no homeomorphism that transforms the phase space portrait for a time-irreversible equation into a phase space portrait for a time-reversible equation. Hence, the Boltzmann equation is structurally unstable. Moreover, the law family for the Boltzmann equation is structurally unstable as a family, for reasons similar to those given in the case about phase transitions. Hence, since systems that exist in the Boltzmann-Grad limit do not qualitatively approximate systems that do not exist in this limit, RS's account of idealized explanation does not show how the Boltzmannian account of irreversibility is explanatory despite being idealized.

Finally, consider Kitcher's account. According to Kitcher, an explanation that involves an idealized description must involve a prologue and an epilogue. The epilogue for the Boltzmannian account of irreversibility must show the ways in which real systems differ from idealized systems that exist in the Boltzmann-Grad limit.

One of the ways these systems differ, according to the Boltzmannian account and the recurrence paradox, is that real systems do not exhibit irreversible behavior. Hence, the epilogue for the Boltzmannian account cannot show how irreversibility in real systems is different from irreversibility in the idealized system, because, according to the Boltzmannian account, there is no irreversibility in real systems (that do not exist in the Boltzmann-Grad limit). There is only irreversibility in systems that exist in the Boltzmann-Grad limit. Hence, Kitcher's account of idealized explanation does not show how the Boltzmannian account of irreversibility is explanatory despite being idealized.

This is not a surprising result, considering that there is no justification, in the sense Kitcher requires, for the prologue of the Boltzmannian explanation. For the idealization of the Boltzmann-Grad limit has a high probability of making a significant difference to the phenomenon of interest, the irreversible approach of non-equilibrium systems to states of equilibrium. (As before, this shortcoming of Kitcher's account is not due to the particular conditions he requires of the prologue and epilogue.)

3.5 The Paradox of Ineliminable Idealization

Any philosophical account of how idealized descriptions can be explanatory despite being false must require idealized descriptions to bear an appropriate relation to correct descriptions in order to be explanatory. The extant accounts of idealized explanation by Laymon, Rueger and Sharp, and Kitcher provide three different conditions which, when satisfied by an idealized description, allow that description to be explanatory. None of these conditions is satisfied by the statistical mechanical explanation of phase transitions or by the Boltzmannian explanation of irreversibility.

The common flaw with these accounts of idealized explanation is that they are unable to accommodate explanations that require, in principle, an appeal to idealization. There is no requisite modal auxiliary for an idealized description of the explanandum phenomenon in such an explanation, because making the description completely realistic (non-idealized) results in a failure to describe the phenomenon. Nor is an idealized description of the explanandum phenomenon in such an explanation structurally stable in any relevant sense, because the behaviors exhibited by the non-idealized systems are qualitatively different from the behaviors exhibited by their idealized versions. Likewise, there is no epilogue to show how the phenomenon in the non-idealized system differs from the phenomenon as it occurs in an idealized version of such a system, because the phenomenon does not obtain in the non-idealized system. (Note that this common flaw is not due to treating idealizations as distortions.)

Any account that interprets idealizations as distortions – and thereby purports to show how idealized descriptions can be explanatory despite being false – will be unable to accommodate explanations of phenomena that can only be described by appeal to idealization, owing to what I call the paradox of ineliminable idealization. This paradox is a schema for generating inconsistencies. Before presenting the schema in its generality, I present an instantiation, as it appears in the statistical mechanical account of phase transitions.

1. Every real system has only finitely many particles.

2. Some real systems undergo phase transitions.

3. A system undergoes a phase transition only if the system is in the thermodynamic limit.

4. Any system in the thermodynamic limit is a system that has infinitely many particles.

These claims are jointly inconsistent.

I call this the *paradox* of phase transitions, because it a set of very plausible assumptions that cannot all be true. The first claim is ontological, agreed upon by all parties. The second claim is the explanandum. The third claim is a mathematical result from statistical mechanics, a property of the Helmholtz free energy per particle. For, according to statistical mechanics, a system undergoes a phase transition only if the Helmholtz free energy per particle for the system develops a singularity; and the Helmholtz free energy per particle develops singularities only in the thermodynamic limit. The fourth claim is justified by the interpretation of the thermodynamic limit as a distorting idealization. The thermodynamic limit is an idealization, because it replaces one description of a system with a description that is, in some sense, simpler. And if this idealization is a distortion, then the syntax '$\lim N \to \infty$' of the thermodynamic limit means 'limit in which the system has infinitely many particles' – i.e., if the limit is a distorting idealization, then systems in the limit have infinitely many particles.

The paradox of phase transitions generalizes. Some form of this paradox is present for any phenomenon that requires, for its explanation, appeal to an idealization. For any such phenomenon, it is possible to construct an argument of the following form:

1. F is some property of every real system (or every real system of interest).

2. P is some phenomenon that occurs in some real systems.

3. P occurs only in systems in an idealizing limit I that idealizes F.

4. Any system in I has a property that real systems lack (and does not have the property F).

These suppositions are jointly inconsistent. I call this the paradox of ineliminable idealization.[53]

The Boltzmannian account of irreversibility fits this form. To see this, let F be the property of having only a finite number of particles in the system. Let P be an irreversible, finite-time, monotonic approach to equilibrium by a system initially not in equilibrium; I call this 'exhibiting irreversibility'. These qualifications on P are intended to ensure that the Boltzmannian approach is the only (currently known) approach that seeks to explain P. The idealizing limit I is the Boltzmann-Grad limit. Then the following argument results:

1. Every real system has only finitely many particles.

2. Some real systems exhibit irreversibility.

3. A system exhibits irreversibility only if the system is in the Boltzmann-Grad limit.

4. Any system in the Boltzmann-Grad limit is a system that has infinitely many particles.

These claims are jointly inconsistent. The first claim is ontological, agreed upon by all parties. The second claim is the explanandum. The third claim is a result of accommodating the Boltzmann equation to the recurrence objection and Poincaré's

[53]The paradox of phase transitions fits this form: let F be the property of having only a finite number of particles in the system; let P be the occurrence of a phase transition; the idealizing limit I is, of course, the thermodynamic limit.

theorem. The fourth claim is justified by the interpretation of the Boltzmann-Grad limit as a distorting idealization. The limit is an idealization, because it replaces one description of a system with a description that is, in some sense, simpler. And if this idealization is a distortion, then the syntax '$\lim N \to \infty$' of the Boltzmann-Grad limit means 'limit in which the system has infinitely many particles' – i.e., if the limit is a distorting idealization, then systems in the Boltzmann-Grad limit have infinitely many particles.

The source of all these paradoxes is the interpretation of idealizations as distortions. It is an empirical issue whether there are phase transitions or systems that exhibit irreversibility; the existence of such phenomena should not be ruled out *a priori*, nor even on the basis of an optional interpretation of idealizations (if there are alternative interpretations, that is). In order to accommodate the existence of ineliminably idealized explanations and avoid the paradox of ineliminable idealization, it must be possible for systems in the relevant idealizing limit to have the properties that are idealized by the limit; and this requires discarding the interpretation of idealizations as distortions.[54] Consequently, since some idealized explanations are ineliminably idealized (explanations like the statistical mechanical account of phase transitions and the Boltzmannian account of irreversibility) idealizations are not distortions – or, at least the thermodynamic limit and Boltzmann-Grad limit are not distortions.

3.5.1 The Role of Interpretation

One might suppose that the interpretation of idealizations as distortions is not necessary for motivating the paradox of irreversibility (or any other version of the paradox

[54] I presume that the first three premises of each paradox are true.

of ineliminable idealization, for that matter), on the grounds that the premise justified by this interpretation is superfluous. According to this way of reasoning, the first three premises of the paradox are already inconsistent: the first two entail that some systems with finitely many particles exhibit irreversibility, while the third entails that only systems with infinitely many particles exhibit irreversibility. Hence, the fourth premise is unnecessary; it need not be justified by the interpretation of the Boltzmann-Grad limit as a distortion. Therefore, according to this line of thought, interpreting idealizations as something other than distortions does not avoid the paradox, because the cogency of the paradox is independent of how idealizations are interpreted.

This way of reasoning assumes that the syntax '$\lim N \to \infty$' (part of the Boltzmann-Grad limit) automatically means 'limit in which the system has infinitely many particles', because it assumes that systems in the Boltzmann-Grad limit have infinitely many particles. This assumption tacitly requires that the Boltzmann-Grad limit be interpreted as a distortion. If the Boltzmann-Grad limit attributes to a real system a property that the real system does not have, then (by definition) the idealizing limit is a distortion. If the syntax '$\lim N \to \infty$' automatically means 'limit in which the system has infinitely many particles', then Boltzmann-Grad limit attributes the property of having infinitely many particles to a real system – it attributes to a real system a property that the real system does not have, because all real systems have only finitely many particles. And this means that the limit is being interpreted as a distortion.

If the Boltzmann-Grad limit is not interpreted as a distortion, the syntax 'lim $N \to \infty$' need not automatically mean 'limit in which the system has infinitely many particles', and systems in the Boltzmann-Grad limit need not have infinitely many particles. This allows for the possibility of interpreting the limit as something other than a distortion, in such a way that the syntax 'lim $N \to \infty$' means something other than 'limit in which the system has infinitely many particles' and in such a way that, given this alternative meaning assigned to the syntax, the Boltzmann-Grad limit does not attribute to real systems a property that real systems do not have. (Providing the details for this possibility is the task of Chapter Four.) Since this sort of interpretation is possible, it is not necessary for systems in the Boltzmann-Grad limit to be systems that do not have only finitely many particles. That is, it is not necessary that systems in the Boltzmann-Grad limit be non-real systems. Hence, the first three premises of the paradox of irreversibility are not inconsistent without the fourth premise. The claim that the syntax 'lim $N \to \infty$' automatically means 'limit in which the system has infinitely many particles' amounts to a tacit endorsement of the fourth premise of the paradox.

3.5.2 A Physicist's Rejoinder

Sometimes the claim is made that the appeal to the thermodynamic limit in the statistical mechanical account of phase transitions is a negligible approximation. This approach – what I'll call the physicist's approach, since it has been given to me by several physicists – takes the thermodynamic limit to be a distortion. The approach also holds that accounts involving distortions can be explanatory if the error due to the distortions is negligible (supposing, for the sake of argument, that there is some non-arbitrary range for what counts as negligible). As regards the statistical mechanical account of phase transitions, the error due to the thermodynamic limit is taken to be unproblematic. Axel Gelfert expresses the general line of reasoning:

> One might worry that the qualitative difference between a *finite* and an *infinite* system could not be greater and, hence, that the thermodynamic limit would necessarily be a wild extrapolation indeed, but given the number of particles in a macroscopic system, typically of the order $N \sim 10^{23}$, and the statistical result that the (relative) error of a statistical average behaves as $\sim 1/\sqrt{N}$, the expected accuracy of the approximation can be seen to be more than satisfactory for most experimental and theoretical purposes ([37], p. 4).

As a distortion, the thermodynamic limit takes the number of particles in the system to infinity. In reality the number of particles in the system is of order 10^{23}; this is very large but not infinite. The relative error of a statistical average is approximately $1/\sqrt{N}$. The statistical mechanical account uses $N \to \infty$ rather than $N = 10^{23}$. Hence, according to the physicist's approach, the error of idealizing the system to be infinite in size is negligible – on the order of 10^{-12} or 10^{-13}. Even though the account of phase transitions appeals to a falsehood, the error due to that falsehood is negligible and, accordingly, does not prevent the account from being explanatory.

This strategy, whatever its merits in other cases, is unsuccessful in the case of phase transitions. Phase transitions occur *only* in the thermodynamic limit. If that limit is a distortion, then phase transitions occur only in systems that have infinitely many particles; if the number of particles is "very large" but not *arbitrarily* large (infinite), no phase transitions occur (according to statistical mechanics). So, even if there is a statistical sense in which the error due to the thermodynamic limit is negligible, there is a more fundamental sense in which the error due to the thermodynamic limit is not negligible, because the thermodynamic limit is ineliminable.

Without the “error” introduced by an appeal to the thermodynamic limit, statistical mechanics fails to describe the occurrence of phase transitions in real systems. Even if measurements never notice the difference between a system with infinitely many particles and a system with 10^{23} particles, there is a well-defined theoretical difference. The difference is that between a system that *can* undergo phase transitions and one that *cannot*. It is this theoretical difference, emphasized by the paradox of ineliminable idealization, that prevents statistical mechanics from explaining phase transitions without appealing to the distortion of the thermodynamic limit.

Of course, one might respond to this kind of argument with the contention that the statistical mechanical account is after the wrong explanandum. One might hold, for example, that phase transitions should not be thought of as singularities, perhaps on the grounds that “most physicists would not expect to be able to measure ‘singularities’ in the first place, as these will always be smoothed out one way or another” ([37], p. 10). A discussion of this sort of position awaits a later chapter.

3.6 Conclusion

Conservatively speaking, the paradox of ineliminable idealization shows that some idealizations are not distortions, namely, the ineliminable idealizations that occur in ineliminably idealized explanations (for example, the thermodynamic limit and the Boltzmann-Grad limit). In the absence of an independent, principled reason for interpreting some idealizations as distortions but not others, it is *ad hoc* to limit the conclusion of the paradox to the claim that only some idealizations are not distortions. A uniform interpretation of idealizations is preferable to a non-uniform one, if a uniform interpretation is possible. (This argument is pursued further in the next chapter.)

With a promissory note that such an interpretation is possible, the interpretation of idealizations as distortions is rejected. This removes the necessity of adopting an account of idealized explanation that allows idealized descriptions to be explanatory despite being false, since idealized descriptions need not be false if idealizations are not distortions. So: if idealizations are not distortions, what are they? An interpretation which allows idealized descriptions to be true of real systems – and thereby avoids the paradox of ineliminable idealization – is in order.

Although an alternative interpretation of idealization is necessary for an adequate philosophical account of why the statistical mechanical account of phase transitions and the Boltzmannian account of irreversibility are explanatory, it is not sufficient. An adequate account must also accommodate explanations that can only be provided by appeal to idealization. An adequate account must not only re-interpret idealizations, but also show how accounts that appeal to idealizations, so interpreted, can be explanatory even when the appeal to idealizations is ineliminable. Another task of the next chapter is to develop such an account.

CHAPTER 4

IDEALIZATIONS AS ABSTRACTIONS

Some idealized explanations are ineliminably idealized – they require appeal to an idealization. The statistical mechanical explanation of the occurrence of phase transitions requires an appeal to the thermodynamic limit; and the Boltzmannian explanation of irreversibility requires an appeal to the Boltzmann-Grad limit. These accounts differ from paradigmatic cases of idealized explanation, because paradigmatic cases are not ineliminably idealized. For instance, the simple pendulum provides an explanation of the rough proportionality between a pendulum's period and the distance between its pivot point and center of mass. The simple pendulum is an idealized version of real pendula that, among other things, idealizes the medium in which real pendula oscillate and the friction that real pendula have at their pivots. The proportionality between a pendulum's period and the distance between its pivot point and center of mass can be explained without appealing to any of these idealizations; the explanation of this proportionality is not ineliminably idealized.

The paradox of ineliminable idealization shows that the existence of ineliminably idealized explanations is incompatible with the interpretation of idealizations as distortions (see Chapter Three). For instance, a key idealization that occurs in the explanations of phase transitions and irreversibility is the limit in which a system's particle number $N \to \infty$. This limiting idealization is ineliminable, in principle, from these explanations: there can be no *explanation* of phase transitions or irreversibility without appealing to the limit in which $N \to \infty$, because there can be no *description* of phase transitions or irreversibility without appealing to this limit. If this limit is a distortion, it is the limit in which the number of particles in a system becomes infinite; and the systems in which phase transitions and irreversible behavior occur are systems in which the number of particles is infinite. No real system has infinitely many particles, however. Hence, if the $N \to \infty$ limit is a distortion, we cannot explain (at least by these methods) the occurrence of phase transitions and irreversibility in real systems.

The existence of ineliminably idealized explanations, such as the explanations of phase transitions and irreversibility, cannot be accommodated merely by allowing explanations in which the explanans is false. If idealizations are distortions, then the paradox of ineliminable idealization shows that the explanandum for any ineliminably idealized account is false of real systems.[55] For instance, if idealizations are distortions, the paradox of irreversibility shows that real systems do not exhibit irreversible behavior; and if real systems do not exhibit irreversible behavior, there can be no explanation of why real systems exhibit irreversible behavior. Generally, if an explanandum is false, there can be no explanation of why the phenomenon it describes obtains. Hence, if idealizations are distortions, then for each ineliminably idealized account that purports to be an explanation, there is a version of the paradox of ineliminable idealization showing that the account is not explanatory because

[55]Taking the paradox to entail this result presumes that every real system has only finitely many particles and that the explanandum phenomenon occurs only in systems that are in a limit that idealizes the particle number of a system. I adopt these presumptions henceforth; the first is obviously correct, and Chapter Three contains arguments for the second.

its explanandum is false. Merely allowing explanations in which the explanans can be false does not accommodate the existence of ineliminably idealized explanations, because if idealizations are distortions then the explanandum of an ineliminably idealized account is false and an account with a false explanandum cannot explain why that explanandum is true.

Prima facie, the accounts of phase transitions and irreversibility are explanatory. For instance, the accounts are appropriately law-like and invoke well-established argument patterns (for details, see Chapter Three). So the existence of ineliminably idealized explanations should not be denied merely on the basis of the paradox of ineliminable idealization. And it need not be denied if it is possible to interpret idealizations as something other than distortions. The aim of this chapter is to provide an account of idealized explanation that involves an alternative interpretation of idealizations. This account is intended not only to be compatible with the existence of ineliminably idealized explanations, but also to show why the statistical mechanical account of phase transitions and the Boltzmannian account of irreversibility are explanatory despite being ineliminably idealized.

The chapter divides into three parts. The first part develops an interpretation of idealizations as abstractions, distinguishing this interpretation from the one that treats idealizations as distortions and showing that this alternative interpretation does not interfere with the mathematical roles of idealizations. The second part of the chapter develops an account of idealized explanation that treats idealizations as abstractions; according to this account, idealized explanations turn out to be a special kind of incomplete explanation. The third and final part of the chapter shows how this account of idealized explanation accommodates the ineliminably idealized explanations of phase transitions and irreversibility.

4.1 An Alternative Interpretation

The aim of this section is to develop an interpretation of idealizations as abstractions. The section distinguishes abstractions from distortions, and subsequently provides an alternative to the interpretation of idealizations as distortions.

4.1.1 Distortions vs. Abstractions

There is a common distinction within the philosophical literature on idealization, between distortions and what are called abstractions. May Brodbeck describes this as "a difference between abstraction and falsification [i.e., distortion], between not saying everything and saying what is not so" ([13], p. 460). Onora O'Neill concurs, holding that "We abstract whenever we [do something] on a basis that *brackets* some predicates, that is indifferent to their satisfaction or non-satisfaction," while we distort whenever we deny those predicates and assert their absence, or else assert that absent predicates obtain ([89], pp. 67-68). According to Ernan McMullin, idealization "may involve a distortion of the original [system, description, etc] or it can simply mean a leaving aside of some components in a complex in order to focus the better on the remaining ones" ([79], p. 248). This distinction, between distortions and abstractions, is a distinction between falsehoods and omissions: distortions falsify, whereas abstractions omit and need not falsify.

Consider a familiar idealized system, interpreting its characterizing idealizations first as distortions and then as abstractions in order to illustrate the difference between these interpretations. A damped simple pendulum is a pendulum that, among other things, is only subject to forces due to gravity and the damping of its surrounding medium (e.g., air). The damping tends to make the pendulum stop its oscillations. The behavior of a damped simple pendulum is correctly described by the following equation:

$$\ddot{\theta} + b\dot{\theta} + \frac{g}{L}\sin\theta = 0,$$

where θ is the angular displacement of the pendulum (this is a function of time), L is the distance from the pivot of the pendulum to its bob, g is the strength of the gravitational force, and b is the strength of damping.

In the idealized $b \to 0$ limit, the equation for the damped simple pendulum reduces to the equation for the simple pendulum, which is given as:

$$\ddot{\theta} + \frac{g}{L}\sin\theta = 0.$$

The mathematical role of the $b \to 0$ idealization is to transform the equation for the damped simple pendulum into the equation for the simple pendulum. This equation is to be understood as characterizing a pendulum in the limit where the amount of damping $b \to 0$.[56] There are (at least) two ways to interpret what the equation for the simple pendulum characterizes, one for each way of interpreting the $b \to 0$ idealization.

First, one might interpret the $b \to 0$ limit as a distorting idealization. Under this interpretation, the idealization says that the amount of damping on the pendulum is arbitrarily close to zero. And the equation for the simple pendulum characterizes a

[56]The equation itself does not contain a term for damping; so this way of understanding the equation cannot be "read off" the equation itself. Nonetheless, the equation is about something. And what the equation is about, in part, is a pendulum for which the amount of damping $b \to 0$. The equation is also about a pendulum in which the friction at the pivot $F_f \to 0$, among other things. But I ignore these further complications, to keep the discussion simple and because they are not salient to the purpose of the example.

pendulum subject to a vanishingly small amount of damping. Even if the equation for the damped simple pendulum is true of some real pendula, the idealized equation for the simple pendulum, so interpreted, is false of all real pendula.

Second, one might interpret the $b \to 0$ limit as an abstracting idealization. Under this interpretation, the idealization says that the amount of damping on the pendulum is to be ignored (rather than made to be arbitrarily small); and the equation for the simple pendulum provides a partial characterization of the damped simple pendulum, a characterization that ignores the amount of damping on the pendulum. Whereas the distortion-interpretation of the $b \to 0$ limit incorrectly represents the amount of damping on the damped simple pendulum by (incorrectly) attributing a vanishingly small amount of damping to the pendulum system, the abstraction-interpretation of the same limit fails to represent the amount of damping on the pendulum by ignoring this feature of pendulum systems. Under this interpretation, the idealization does not attribute an incorrect amount of damping to the pendulum. Nor does it say that there is a non-zero amount of damping on the pendulum, since ignoring the amount of damping is consistent with the pendulum having a zero amount of damping (e.g., swinging in a vacuum). As an abstraction, the idealization of damping simply does not specify the amount of damping on the pendulum; and in particular it does not specify an incorrect amount of damping.

This example supports a more concise characterization of abstraction. According to Anjan Chakravartty, abstraction is "a process whereby only some of the potentially many relevant factors or parameters present in reality are built-in to a model concerned with a particular class of phenomena" ([23], p. 327). According to Margaret Morrison, an abstract description is one that "does not include all of the systems

[sic] properties, leaving out features that the systems [sic] has in its concrete form" ([85], p. 38 fn. 1). If idealizations are abstractions, then an idealization replaces one description of a system with a simpler description that fails to attribute to the system at least one feature that the system has, without thereby attributing to the system a feature it does not have. The resultant abstract description ignores some feature of the system; and the idealizations used to obtain this description function as "inference tickets" that transform one description into a less complete description. Hence, if idealizations are abstractions, an idealized description of a system is a partial (incomplete) description of that system.

The term "feature" in this characterization of abstractions is intended to provide a quick way of referring to the value or amount of some property. In many cases, ignoring the amount of some property effectively amounts to ignoring the property itself. For example, sometimes ignoring the amount of mass of a particle amounts to ignoring that the particle has mass at all. So sometimes I will speak as if a feature is a property of the system itself. This allows me to follow others in speaking of abstractions as ignoring properties of systems rather than amounts of those properties; but it also allows me to expand their notion of abstraction to cover cases in which ignoring the amount of some property does not amount to ignoring the property itself. For example, consider the idealizing limit in which the particle mass $m \to 0$ for each particle in a system and the system's particle number $N \to \infty$ while the system's total mass $M = mN$ remains finite and non-zero. As an abstraction, this idealization ignores the amount of mass for every particle in the system but does not ignore the amount of mass of the system itself. And since the system having some mass entails that at least some of the system's components have mass, this idealization does not ignore the fact that at least some particles of the system have the property of mass.

The interpretation of idealizations as abstractions does not entail that every abstraction is an idealization. Some abstractions are not idealizations, because not all abstractions replace one description of a system with a simpler description of the same system. Anatomy and physiology textbooks routinely discuss the different systems of the body in abstraction from other systems of the body. For instance, discussions and illustrations of the skeletal system often ignore the cardiovascular and nervous systems. Yet the descriptions of the skeletal system that ignore other bodily systems are not simpler, in any computational sense, than a unified, complete description of all of the body's systems.

The interpretation of idealizations as abstractions rather than distortions does not interfere with the mathematical role of idealizations. Consider again the example of the damped simple pendulum. Under both interpretations, the simple pendulum has the same phase space portrait. However, the interpretation of what the phase space portrait represents depends upon the interpretation of the $b \to 0$ limit. If the limit is a distortion, the phase space portrait represents the trajectory of a simple pendulum subject to a vanishingly small amount of damping. If the limit is an abstraction, the same phase space portrait only partially represents the trajectory of a damped simple pendulum, by ignoring the amount of damping on the pendulum. (This is similar to the way a stick figure only partially describes a person's appearance.) Hence, the $b \to 0$ limit can be interpreted in at least two ways, without interfering with the mathematical role of the idealization. This result is expected to generalize to other idealizations.

Attention to a formal definition of limits further supports the mathematical legitimacy of treating limiting idealizations as abstractions. It is typical to define the limit of a function at a point in the following manner: $\lim_{x \to c} f(x) = L$ iff: $(\forall \varepsilon > 0)(\exists \delta > 0)$ such that $(\forall x)$ if $0 < |x - c| < \delta$, then $|f(x) - L| < \epsilon$. This does not define what it is for one *function* to be a limit of some other function (because L is a point, not a function). But the complications involved in formulating such a definition (e.g., replacing $f(x)$ with a multivariable function $f(x, y)$ and L with a function $g(x)$, defining a measure for the norm $\|f(x, y) - g(x)\|$, and so on) are incidental to whether it makes sense, from a mathematical perspective, to treat a limiting idealization of the form "$x \to c$" as an abstraction: if such an interpretation is legitimate when the limit of a function is at a point, it should also be legitimate when the limit of a function is another (simpler) function.

Suppose, then, that the limit $x \to c$ is a limiting idealization of some non-idealized system S. Treated as a distortion, this limit says that there is a value for the property of S represented by the physical magnitude x and that the value of x is arbitrarily close to c;[57] but this is false, because the actual value of x for any particular non-idealized system S is not arbitrarily close to c. In contrast, as an abstraction the limit says that there is some value for the property of S represented by x, and that's it; the idealization does not specify what that value is, nor does it say that the value is non-zero or close to c. As an abstraction, the idealization is not false – provided, of course, that there is some value for the property in S represented by x.

[57]More correctly, the limit says that the value of x is bounded around c and that this bound can be made to be arbitrarily small. But saying that the value of x is arbitrarily close to c seems to be a less cumbersome way of speaking.

Moreover, the difference $|x - c|$ that appears in the definition of the limit is well-defined if the limit is treated as an abstraction. This is because that difference is well-defined so long as c has the same dimensions as x, and this condition is satisfied when the value – but not the presence – of x is ignored. (If the idealization $x \to c$ were to ignore x itself rather than the amount of x, it is not clear that systems in the limit $x \to c$ would have some property represented by a magnitude with the same dimensions as x.) Since the only part of the definition of the limit in which the parameters involved in the idealization occur is the expression "$0 < |x - c| < \delta$", and since that expression is well-defined if the limit manages to ignore the value of x (but not x itself), treating the limit $x \to c$ as an abstraction is mathematically legitimate.

4.1.2 Abstract Descriptions

If idealizations are abstractions, then a description of a system obtained by appeal to idealizations is incomplete – it is a description that ignores certain features of the system under consideration. Such descriptions are commonplace. If someone says that the number of coins in his pocket is odd without saying anything else, his description of his pocket's contents is abstract in virtue of leaving aside details about how many coins are in his pocket. And if someone says that the gas inside the tube is a noble gas without saying anything else, her description of the tube's contents is abstract in virtue of leaving aside details about which noble gas is in the tube. None of these abstract descriptions are false.

This way of thinking about ordinary abstract descriptions is applicable to idealized descriptions. Let *f(x,y)* $= 0$ be an equation that is not idealized in any way and that correctly characterizes a physical system S. Let *g(x)* $= 0$ be the equation obtained by taking the idealizing limit of $f(x,y)$ in which y approaches zero:

$$\lim_{y \to 0} f(x, y) = g(x),$$

so that $g(x) = 0$ characterizes an idealized version of S. Then there are two salient ways to understand the relation between the equation $g(x) = 0$ and the system S.

The equation $g(x) = 0$ can be understood as purporting to stand in a correspondence relation to S (or whatever relation $f(x,y) = 0$ bears to S). Since the limit in which y approaches zero is an idealizing limit, however, $g(x) = 0$ fails to stand in such a relation: $g(x) = 0$ is false of S, because the idealization used to obtain $g(x)$ from $f(x,y)$ is false of S. For instance, if $f(x,y) = 0$ is the equation of motion for a damped simple pendulum, if $g(x) = 0$ is the equation of motion for an undamped simple pendulum, and if the limit "$y \to 0$" idealizes the damping on pendula, then $g(x) = 0$ is false of damped simple pendula because such pendula are subject to more than an arbitrarily small amount of damping. This way of understanding the relation between an idealized description and the physical system it purports to characterize results from treating idealizations as distortions.

It is not mandatory to understand the $g(x) = 0$ as purporting to bear a correspondence relation to S but failing to do so. The same equation can be understood as standing in a correspondence relation to an abstract version of S rather than to S itself. Let S^A be this abstract version of S, so that $g(x) = 0$ correctly characterizes S^A. Then the "$y \to 0$" idealization functions to transform one description – viz., $f(x,y) = 0$ – into a (more) incomplete description – viz., $g(x) = 0$; and the "$y \to 0$" idealization determines which details $g(x) = 0$ ignores about S and the respects in which S^A is an abstract version of S. Provided that there is an appropriate relation between S^A and S itself, it is possible for $g(x) = 0$ to be true of S because, as an abstract description of S, $g(x) = 0$ need not be false of S. This point generalizes: an idealized description need not be false if it is an abstract description, because it only purports to characterize real systems indirectly, based upon whether the abstract system it characterizes bears an appropriate relation to real systems.

4.1.3 Incompleteness and Truth

The interpretation of idealizations as abstractions rather than distortions, and the understanding of idealized descriptions as abstract descriptions, have repercussions for the correctness conditions of idealized descriptions. Consider, once more, the equation for the simple pendulum, obtained through appeal to the $b \to 0$ limiting idealization:

$$\ddot{\theta} + \frac{g}{L} \sin \theta = 0.$$

If the $b \to 0$ limit is a distortion, then this equation describes the behavior of a pendulum subject to a vanishingly small amount of damping (among other idealizations). And it is always false of real pendula.

However, if the $b \to 0$ is an abstraction, the equation partially describes the behavior of a pendulum that is subject to damping (among other things). There is a sense in which partial descriptions can be "true" despite being partial; a reasonable assumption is that whether a partial description is true of a system depends upon whether what is ignored is "relevant" to the system.[58] (The notion of "relevance" is

[58]There are alternative suggestions in the literature. For instance, Nancy Cartwright claims that equations that ignore some details of real systems describe capacities of abstract systems, and that when a real (concrete) system has the same capacities as an abstract system, the abstract equation describing that abstract system is also true of the real system. (See [17]; [18].) The suggestion here is more ontologically parsimonious than Cartwright's, since it does not postulate the existence of "capacities". Also, Cartwright holds that a description that is abstract relative to a more concrete set of descriptions never applies unless one of the more concrete descriptions applies ([19], p. 259). The suggestion here, in contrast, allows a description that is abstract relative to a more concrete description to be true of a system even if the more concrete description is not true of the system.

discussed more extensively later in this chapter.) For instance, since the equation for the simple pendulum ignores some features of a real pendulum and sometimes partial descriptions can be "true", a reasonable assumption is that whether the equation for the simple pendulum is true depends upon whether the ignored features are "relevant" to the real pendulum. Relevance is phenomenon-relative: a feature might be relevant with respect to one phenomenon of a system but not with respect to some other phenomenon of the same system, because not every phenomenon of a system always depends upon every feature of the system.

Since the correctness of a partial description depends upon the relevance of what is ignored, and since relevance is phenomenon-relative, it follows that an abstract description is true of a system with respect to a given phenomenon of the system just in case what is idealized is not relevant to that phenomenon in that system. So, for instance, the equation for the simple pendulum is true of a real pendulum with respect to the rough, qualitative proportionality between the pendulum's period and its length just in case what is ignored in idealizing the pendulum (such as the amount of damping on it) is irrelevant to that proportionality. Hence, if idealizations are abstractions rather than distortions, the equation for the simple pendulum can be true of a real pendulum with respect to the rough proportionality between the pendulum's period and its length, even if there is not a vanishingly small amount of damping on the pendulum. This possibility, of idealized equations being true of the systems they characterize, is absent if idealizations are distortions.

4.2 Incomplete Explanation

Having set forth an interpretation of idealizations as abstractions in the previous section, this section provides an account of idealized explanation appropriate to such an interpretation. Since this dissertation is not concerned with explanation itself, but with explanation (whatever that turns out to be) that is idealized, the presentation of this account takes for granted that there is some adequate account of non-idealized explanation, and presents the modifications to be made to such an account so that it accommodates idealized explanations as well. The idea is that there is a set of conditions, Θ, given by some (unspecified) philosophical account of scientific explanation, such that a *non-idealized* scientific account of a phenomenon is a scientific explanation of that phenomenon if the scientific account satisfies Θ; and, furthermore, that an *idealized* scientific account of a phenomenon is explanatory if it satisfies both the conditions Θ and *additional* conditions that pertain specifically to *idealized* explanations. The task in this section of the chapter is to provide these additional conditions, the conditions that pertain specifically to explanations that are idealized. These additional conditions will be compatible with a variety of specific proposals for the content of Θ.[59]

[59]Alexander Bird argues that it is impossible to give a model template such that all and only accounts that satisfy that template qualify as explanations ([9]). Even if Bird is correct, idealized accounts must satisfy more constraints than non-idealized accounts in order to be explanatory, because the idealized accounts must satisfy all of the constraints that pertain to non-idealized explanation *plus* the constraints the pertain specifically to idealized explanation. Hence, even if it is not possible to provide a set of conditions for non-idealized explanation, it is possible to provide the additional constraints that pertain to idealized explanations without providing an account of non-idealized explanation.

4.2.1 The Relevance Requirement

Any philosophical account of idealized explanation that treats idealizations as abstractions must supplement the account of non-idealized explanation with a criterion that distinguishes explanatory abstract descriptions from non-explanatory ones. In its most general form, this criterion requires abstract descriptions to bear an appropriate relation to the systems they incompletely characterize in order to be explanatory. Cases of non-idealized explanations that involve abstract, and therefore incomplete, descriptions are useful guides to what such a criterion should be. Consider, for instance, the account of the precession of Mercury's perihelion as given by the general theory of relativity. Mercury's orbit around the sun is elliptical, but the ellipse moves with every orbit. More specifically, the point at which Mercury is closest to the sun (the perihelion) rotates around the sun (precesses) at a rate of about 5600 seconds of arc per century (1 second of arc = 1/3600 degrees), as measured from the Earth. This is an interesting phenomenon, and one might very well ask why Mercury's perihelion precesses at about this rate.

Newtonian mechanics predicts that Mercury precesses at a rate of 5557 arcseconds per century.[60] The general theory of relativity predicts an additional 42.9195 seconds of arc per century, which is close to the observed 43.105 seconds of additional arc per century. The equations of general relativity are explanatory, on any reasonable account of what would count as explanatory. Those equations, along with details about some additional properties of the solar system, entail that Mercury's perihelion

[60] This led the astronomer Le Verrier to postulate the existence of a planet, Vulcan, between Mercury and the sun, having features that would account for the additional 43 seconds of arc per century. This is similar to the postulation of the existence of Neptune to account for the orbit of Uranus; the difference is that Vulcan does not exist.

precesses at a rate of about 5600 seconds of arc per century. It is reasonable to conclude that general relativity explains why Mercury's perihelion precesses at a rate of about 5600 arcseconds per century (i.e., at a rate of 5600 $\pm$ 1 arcseconds per century).

Nonetheless, this explanation ignores some details of the solar system (the solar system being the system in which the precession occurs). The description of the solar system involved in this account of Mercury's perihelion is, accordingly, an abstract description. For instance, the description ignores the effects of comets on the precession of Mercury's perihelion; it ignores the composition of Mercury or the sun or the other planets in the solar system, and conditions on the surfaces of the planets. And it need not attend to these additional details, because they are, in some sense, irrelevant to the rate at which Mercury's perihelion precesses.

For a second example of an explanation that involves an incomplete description, consider the less astronomical, but no less fascinating, behavior of bubbles in Guinness (a beer). When poured into a glass, some of the bubbles in Guinness appear to go downwards rather than upwards. (This has been confirmed by high speed cameras.) This is surprising, because the bubbles are lighter than the beer, so one would expect them to rise rather than fall. One might very well ask why this phenomenon occurs. What follows is an account given by Richard Zare and Andrew Alexander.

Consider the beer when it has just been poured into a glass and is starting to settle. The bubbles touching the sides of the glass experience drag, which prevents them from flowing upwards; but the bubbles away from the sides, and especially those near the center, do not experience this drag. Unencumbered by drag, the bubbles near the center of the glass go up, as gases in liquids tend to do. As these bubbles rise, they push and pull the surrounding liquid with them. When the bubbles and liquid reach the surface of the beer, the bubbles escape and the liquid flows away from the center, towards the sides of the glass, since it has nowhere else to go. The current due to this flow gets directed downwards by the sides of the glass, again because there is nowhere else for the liquid to go. The flow of liquid moves down the sides of the glass in waves, taking some of the smaller bubbles (with less buoyancy) touching the sides of the glass down with it.

The key components of this account are: that gases rise in liquids when unencumbered; that bubbles of gas encumbered by drag can remain stationary in liquid; that when a stream of liquid flows up within a container and reaches the surface, the stream disperses away from the center; and that when such a flow reaches the sides of the container, the sides direct the flow downwards. There seems to be an appropriate relationship between these components and the explanandum; for instance, the account cites appropriate causes and laws, and accounts with similar structures can be given for other phenomena (how other liquids flow). It is reasonable to conclude that this account explains the behavior of Guinness bubbles.

Nonetheless, this explanation ignores some details of the glass of Guinness. The description of the glass of Guinness involved in this account of the flow of bubbles in Guinness is, accordingly, an incomplete description. For instance, the description ignores details about the drag on bubbles near the sides of the glass, the specific buoyancies of any of the bubbles, and the rate at which the bubbles in the center flow upwards; it also ignores the chemical composition of the glass, the temperature of the beer and the glass; and the way in which the beer is poured. And it need not attend to these additional details; they might provide a fuller explanation, but they are largely irrelevant to why Guinness exhibits this general sort of behavior.[61]

These cases, and others like them, suggest that an incomplete description of some phenomenon that occurs in a system bears an appropriate explanatory relationship to that system if the description ignores only those facts about the system that are, in some sense, irrelevant to the occurrence of the phenomenon of interest.[62] Although the account of Mercury's perihelion provided by general relativity involves an incomplete description of the solar system, that description attends to all of the facts about the solar system that are relevant to Mercury's orbit. Likewise, although the account of the flow of bubbles in Guinness involves an incomplete description of Guinness, that description attends to all of the facts about Guinness that are relevant to its bubbly behavior.

[61]Some details ignored by this account are relevant to the explanation of why the downward flow of bubbles is easier to see in Guinness as opposed to other liquids. High-speed cameras show that some bubbles near the side of glasses containing fluids as simple as water (or water mixed with some sort of fizzing tablet) also flow downwards. This flow is easier to see in Guinness than in water for three main reasons: there is a high contrast between dark Guinness and lighter, cream-colored bubbles, which makes the bubbles more visible; the bubbles in Guinness are smaller, hence more easily pushed around in the glass; the bubbles in Guinness contain nitrogen, which is less likely to dissolve in liquid, hence less likely to enlarge, hence more likely to stay submerged in the liquid given sufficient surface tension with the sides of the container. (In contrast, the bubbles in soda typically contain carbon dioxide, which more readily dissolves in liquid than nitrogen.)

[62]This requirement of relevance is a similar to Hempel's requirement of maximal specificity (see [43], pp. 397-400). Hempel requires that, in constructing inductive-statistical explanations, one include in the premises any knowledge that is both relevant to the explanandum and in-principle available prior to the occurrence of the explanandum. The requirement here is that, in constructing an idealized explanation, one include in the premises everything that is relevant to the explanandum. Although all deductive-nomological explanations satisfy Hempel's requirement, they need not all satisfy the present requirement, since the explanans of an idealized deductive-nomological account might contain an idealization of some feature that is relevant to the explanandum.

4.2.2 Effective Field Theories

This constraint on when incomplete descriptions can be explanatory is consonant with contemporary attitudes toward effective field theories. The notion of an effective theory is an outgrowth of developments in high energy physics, the study of the structure of the atom (nucleus plus surrounding electrons) and the structure of the nucleus. Stephan Hartmann provides a succinct presentation of an effective theory known as the Euler-Heisenberg theory (see [41], pp. 270-273). Here I quickly summarize Hartmann's discussion.

The Euler-Heisenberg theory provides an explanation of photon-photon scattering, which is a process in which two photons scatter and create an electron-positron pair, which then decays into two photons. For high photon energies, photon-photon scattering results in the creation of real electrons and positrons. The Euler-Heisenberg theory, however, focuses on photon-photon interactions in which the photons have energies insufficient for the creation of electrons and positrons. This allows the theory to ignore effects due to electrons and thereby yield an equation that is the basis for many calculations.

The Euler-Heisenberg theory is an *effective* theory, because it only takes into account the photon field, ignoring the electron field. Despite this omission, the theory is valid at energy scales below the threshold for electron production. The reason the theory is valid at a low enough energy scale is that the electron field is *irrelevant* to photon interactions at that scale, and what is true of the abstract version of the real atomic systems described by the Euler-Heisenberg theory is true of the real systems themselves. At higher energies, however, the electron field becomes relevant to photon-photon interactions, and the Euler-Heisenberg theory fails. This is because, at higher energies, photon-photon scattering creates both positrons and electrons, something not predicted by the Euler-Heisenberg theory.

4.2.3 An Account of Relevance

There is an intuitive sense in which some facts about a system are relevant to a phenomenon that occurs in the system and other facts about the system are irrelevant to that phenomenon. For instance, the shape of Mercury's surface is irrelevant to the fact that Mercury's perihelion precesses at a rate of about 5600 arcseconds per century. The composition of a glass of Guinness is irrelevant to the fact that some bubbles in the beer near the side of the glass fall downwards. The electric field is irrelevant to photon-photon interactions at low photon energies. In contrast, the gravitational attraction between Mercury and other planets in the solar system is relevant to Mercury's orbit, and the drag on bubbles touching the sides of a glass of Guinness is relevant to the direction in which those bubbles move.

There is an important difference between details that are ignored by the Euler-Heisenberg theory in its explanation of photon-photon scattering and details that are ignored by, say, General Relativity in its explanation of the precession of Mercury's perihelion. The Euler-Heisenberg theory ignores effects due to the electric field on photon interactions; and the theory itself determines when the electric field is irrelevant to photon interactions. The derivation of the Euler-Heisenberg theory presupposes that the photon energy is small compared to the rest mass of electrons; and various symmetry considerations and dimensional analyses show that when the photon energy is small compared to the rest mass of electrons, the effects of the electron field on photon-photon interactions are suppressed – the electric field is irrelevant to the interactions (see [41], p. 274).

In contrast, General Relativity ignores various details of the solar system on the orbit of Mercury; but the theory does not determine when these details are irrelevant to the rate at which Mercury's perihelion processes. The aim of this subsection is to provide a working account of how to tell whether something is relevant to a phenomenon, when there is no guidance from the theory of that phenomenon.[63] Of course, figuring out whether the conditions provided by this account obtain takes a lot of empirical footwork – that's a job for the scientists. From a philosophical perspective, the aim of giving an account of relevance is to elucidate what one would need to show in order to establish the relevance of something to a given phenomenon. (The reader is welcome to substitute a better account. The aim here is not to defend the correctness of a particular account of relevance, but rather to suggest a plausible account that at least shows the notion of 'relevance' to be non-vacuous.)

Let a *nomic web* for a phenomenon be the set of events, initial and boundary conditions for the system in which the phenomenon occurs, as well as the laws governing that system.[64] Also, let a nomic web for a system in which some phenomenon occurs be the nomic web for that phenomenon. The set of all events, initial and boundary

[63]The kind of relevance to be discussed here is not the same kind of relevance prominent in discussions of asymptotic reasoning. Asymptotic analysis might show that something counted as relevant by the account given below is not relevant in some other, stricter sense. For instance, the only properties that are relevant – in an asymptotic sense of "relevant" – to the universality of critical phenomena are the spatial dimension of a system, symmetry properties of its Hamiltonian, and the fact that the forces between its components are short-range; the microstructural details of the system are largely irrelevant (in the asymptotic sense): see [5], p. 24.

[64]Examples of events: the activation of a photocell by a pulse of light, a sneeze, a baseball colliding with a window. Examples of initial and boundary conditions: a system being in a cylindrical container or having a certain mass. Examples of laws: Newton's law of gravitation, Maxwell's laws of electromagnetism.

conditions, and laws is a nomic web for every phenomenon that occurs in the universe; but most phenomena also have nomic webs that are smaller than the aforementioned set.

Let a model of a nomic web – a *nomic model*, for short – be a linguistic representation of events, initial and boundary conditions, and laws.[65] (If laws are statements, then the linguistic representation of a law is the law itself.) A linguistic representation of a nomic web for a system in which a phenomenon occurs is a nomic model for that phenomenon. But a nomic model for a phenomenon need not be a linguistic representation of the nomic *web* for the phenomenon, because a nomic model for a phenomenon might represent only a *part* of the nomic web for the phenomenon. A nomic model will be said to be *veridical* for a system just in case the statements in the nomic model are true of the system. When a nomic model is veridical for a system, the model is a veridical nomic model for the system.

A set of events, initial and boundary conditions, and laws will be said to *nomically produce* a phenomenon just in case a correct linguistic representation of those events, initial and boundary conditions, and laws (non-circularly) entails that the phenomenon occurs.[66] That is: a set of events, initial and boundary conditions, and laws nomically produces a phenomenon just in case a veridical nomic model for those events, initial and boundary conditions, and laws (non-circularly) entails that the phenomenon occurs.

Let a nomic model be called *deterministic* for a phenomenon that occurs in a system just in case the nomic model (non-circularly) entails that the phenomenon

[65] Not every true conditional belongs to a model of a nomic web; only the lawful conditionals belong.

[66] Why non-circular entailment? To prevent the inclusion, in the nomic model for the phenomenon, of a statement that the phenomenon occurs.

occurs. When a nomic model is deterministic for a phenomenon that occurs in a system, the model will be called a deterministic nomic model for the phenomenon in that system. When the statements in such a model are true of the system, the model will be said to be a *veridical deterministic nomic model* for the phenomenon in that system.

Let a nomic model be called *indeterministic* for a phenomenon that occurs in a system just in case the nomic model (non-circularly) entails that there is a non-zero probability for the occurrence of the phenomenon and both (1) the nomic model is consistent with the truth of the claim that the phenomenon itself occurs in the system and (2) the nomic model is consistent with the truth of the claim that the phenomenon itself does not occur in the system. When a nomic model is indeterministic for a phenomenon that occurs in a system, the model will be said to be an indeterministic nomic model for the phenomenon in that system. When the statements in such a model are true of the system, the model will be said to be a *veridical indeterministic nomic model* for the phenomenon in that system.

These notions are intended to contain no explicit reference to causal notions. Whether there is an implicit reference to causal notions depends upon one's account of causation. For instance, according to the Hempelian account of causation, an event e_1 is the cause of a distinct event e_2 if and only if the statement that e_2 occurs is deducible from the statement that e_1 occurs, laws of nature, and statements that describe appropriate initial and boundary conditions. According to this Hempelian account, a nomic web for a phenomenon is also a causal web for that phenomenon, because the nomic web contains the (putative) causes for the phenomenon.

With these notions in place, it is possible to provide an account of how to tell whether a property is relevant to a phenomenon. More precisely, it is possible to provide two such accounts – one for deterministic phenomena and one for indeterministic phenomena. The account for deterministic phenomena is given first, followed by an illustration of this account and a modification of the account into an account for indeterministic phenomena.[67] (Indeterministic phenomena are "chancy" phenomena, like radioactive decay: think of quantum mechanics.)

According to the account of relevance for deterministic phenomena, the following two-step procedure suffices to determine whether a property (event, initial or boundary condition, law) of a system is relevant to a deterministic phenomenon that occurs in the system.

1. Consider a nomic web for the system in which the deterministic phenomenon of interest occurs. Find a part of this web that suffices to nomically produce the phenomenon of interest. A correct linguistic representation of this part of the nomic web is a veridical deterministic nomic model Θ for the phenomenon of interest.

2. Find a subset Δ of this veridical deterministic nomic model Θ for the phenomenon of interest such that: Δ suffices to nomically produce the phenomenon of interest, and there is no proper subset of Δ that suffices to nomically produce the phenomenon of interest.[68]

[67] The account for deterministic phenomena is adapted from an account given by Michael Strevens ([113]). Strevens's account uses causal notions; and he doesn't bother to say what he means by the causal notions he uses. Strevens' account also does not apply to indeterministic phenomena.

[68] Representing the phenomenon of interest as ψ, the sequent $\Delta : \psi$ should be perfectly valid in Neil Tennant's sense (see [118]).

Call the subset Δ of statements a *model of a minimal kernel* for the phenomenon of interest. The events, initial and boundary conditions, and laws represented in a model of a minimal kernel for a phenomenon are said to be a *minimal kernel* for that phenomenon.[69] And a property (event, initial or boundary condition, law) is *relevant* to a deterministic phenomenon if and only if that property appears in at least one minimal kernel for the phenomenon. (Of course, it takes extensive empirical footwork to determine whether a property appears in at least one minimal kernel for a phenomenon; this is not something to be settled from the armchair.)

Consider the example about the precession of Mercury's perihelion, as a way to illustrate this account of relevance. Let the phenomenon of interest be the precession of Mercury's perihelion at a rate of about 5600 seconds of arc per century. The system in which this phenomenon occurs is the solar system. (It also occurs in larger systems, such as the universe.) The nomic web for the solar system (and for the phenomena that occur in the solar system) includes its size; the number of planets in it; the masses, shapes, and various other properties of each of these planets and their moons; the objects on these planets and moons and the properties of such terrestrial objects; the various comets and asteroids that travel within the solar system and their various properties; properties of the sun; laws about gravity and other forces between the celestial objects; laws about forces between the terrestrial objects on these celestial objects (such as magnetism and electricity); more specific laws about the orbits of the planets; a law about the speed of light; laws from general relativity; and so on.

[69] It is important to note that a minimal kernel (or model thereof) for a phenomenon is not necessarily an explanans for that phenomenon.

Part of this nomic web suffices to nomically produce the precession of Mercury's perihelion at a rate of about 5600 seconds of arc per century. For simplicity, take this part to be the part of the nomic web for the solar system that is explicitly mentioned in the preceding paragraph, with appropriate details filled in (such as the mass and size of each planet, etc). For the sake of illustration, suppose that these statements about the details of the solar system are true; and suppose that these statements (collectively and non-circularly) entail that the perihelion of Mercury precesses at a rate of about 5600 seconds of arc per century.[70] This completes Step 1, yielding a veridical deterministic nomic model for the approximate rate at which the perihelion of Mercury precesses.

Many of the statements in this model are such that, were they to be eliminated from the model, the resultant model would still entail that the perihelion of Mercury precesses at a rate of about 5600 arcseconds per century. These statements include statements about the height and weight of various celebrities, statements about the color of each planet, statements about the precise mass and size of each asteroid in the asteroid belt between Mars and Jupiter; statements about the chemical composition of water; laws about strong and weak nuclear forces; more specific laws about the behaviors of pendula in oil; statements about the composition of various celestial objects; and so on.

Remove these statements, and other like them, from the model that resulted from Step 1, in such a way that (1) the statements remaining in the resultant model (non-circularly) entail that the perihelion of Mercury precesses at a rate of about 5600 arcseconds per century and (2) there is no proper subset of the statements in

[70]It follows that these statements do not include the statement about the precession of Mercury's perihelion.

the resultant model that also entail that the perihelion of Mercury precesses at a rate of about 5600 arcseconds per century. This completes Step 2, yielding a model of a minimal kernel for the approximate rate at which the perihelion of Mercury precesses. The laws and conditions described by this model are a minimal kernel for the approximate rate at which the perihelion of Mercury precesses; they are relevant to the phenomenon of interest.

Of course, some properties that appear in the nomic web for the solar system do not appear in any minimal kernel for the approximate rate at which the perihelion of Mercury precesses. For instance, it is reasonable (for the sake of illustration) to conjecture that the chemical compositions of Mercury and the sun and the other planets in the solar system do not appear in any minimal kernel for the approximate rate at which the perihelion of Mercury precesses. Likewise, it is reasonable to suppose that biological properties of terrestrial species do not appear in any minimal kernel for the approximate rate at which the perihelion of Mercury precesses. Properties like these are irrelevant to the phenomenon of interest.

This account of relevance for deterministic phenomena does not apply to chance phenomena. If the phenomenon of interest is indeterministic, no set of statements about laws, initial and boundary conditions, and events (non-circularly) entails that the phenomenon occurs. For instance, the decay of a particular piece of uranium lacks a veridical deterministic nomic model, because although the occurrence of this decay is consistent with the truths about the world, the non-occurrence of this decay is also consistent with these truths (or so our best sciences say, on their most widely accepted interpretations). Nonetheless, there can be veridical nomic models that entail that there is a chance for such a decay to occur; these are veridical indeterministic nomic models for the phenomenon.

According to the account of relevance for indeterministic phenomena, the following two-step procedure suffices to determine whether a property (event, initial or boundary condition, law) of a system is relevant to an indeterministic phenomenon that occurs in the system.[71]

1. Consider a nomic web for the system in which the indeterministic phenomenon of interest occurs. Find a part of this web that is correctly linguistically represented by a veridical indeterministic model for the phenomenon of interest.

2. Find a subset Δ of this veridical indeterministic nomic model for the phenomenon of interest such that: Δ entails that there is a non-zero probability for the phenomenon of interest to occur, and there is no proper subset of Δ that entails there being a non-zero probability for the phenomenon of interest to occur.

As before, call the subset of statements Δ a *model of a minimal kernel* for the phenomenon of interest. The events, initial and boundary conditions, and laws represented in a model of a minimal kernel for a phenomenon are said to be a *minimal kernel* for that phenomenon. And a property (event, initial or boundary condition, law) is *relevant* to an indeterministic phenomenon if and only if that property appears in at least one minimal kernel for the phenomenon.

[71]The strategy here is to imitate Peter Railton's strategy in adapting Hempel's DN model of explanation to indeterministic phenomena (see [97]).

4.2.4 The Idealization Requirement

In addition to a relevance requirement on idealized explanations, there is an obvious requirement that the explanans of an idealized explanation be idealized in some way: the set of idealizations that occur in the explanans should be non-empty. These two constraints can be added to the (unspecified) conditions that a putative explanans must satisfy in order to be explanatory, yielding a set of conditions that a putative explanans must satisfy in order to be an *idealized* explanation. For instance, according to Salmon, "an explanation of an event involves exhibiting that event as it is embedded in its causal network and/or displaying its causal structure" ([103], p. 325). It is straightforward to modify Salmon's causal account of non-idealized explanation into a causal account of idealized explanation. The modifications involve interpreting idealizations as abstractions, requiring that the event to be explained be exhibited as it occurs in the causal nexus of an idealized version of the system in which the event actually occurs, and requiring that this idealized version of the real system omit only properties that are irrelevant to the occurrence of the event.

Similarly, according to Hempel's DN model of explanation, an explanation is a sound derivation of an explanandum from a set of law-statements and other non-nomological conditions, such that each premise in the derivation is both true and confirmable (empirical) and the inference would be invalid if any law-statement were omitted. Four supplements to the DN model transform it into a model of idealized explanation. First, the model should allow idealizations to occur in the derivation of the explanandum; these idealizations should be interpreted as abstractions, not as distortions. Second, whereas the DN model imposes the condition that, in the absence of appeal to any of laws, the non-nomological conditions by themselves must be insufficient to entail the explanandum, the revised model should impose the condition that, in the absence of appeal to any of laws, the non-nomological conditions and the idealizations used in the account must be insufficient to entail the explanandum. Both conditions are ways of stating that the law-statements must be used in any derivation of the explanandum. Third, the revised model should make the obvious supplementation that the set of idealizations is not empty. For when this set is empty, the explanation is not an *idealized* explanation. Finally, the revised model should impose a requirement of relevance on any idealized descriptions that occur in the explanans.

To illustrate the account of idealized explanation that results from these revisions to Hempel's DN model, consider again the explanation, provided by appeal to the simple pendulum, of the qualitative proportionality between a pendulums's period and its length. The equation of motion for the simple pendulum can be derived from various modeling assumptions about the simple pendulum and Newton's laws of motion. Newton's laws of motion alone do not entail the equation of motion. Modeling assumptions are also required; and many of these assumptions are idealizations. (The details omitted by such idealizations are assumed to be irrelevant to the phenomenon of interest, for the sake of illustration.) From the equation of motion, it is possible to derive an expression relating the period of the simple pendulum to its length; this expression shows that the period is roughly proportional to the pendulum's length. Since the derivation of this result contains idealizations as premises, since it is appropriately law-based (and otherwise satisfies Hempel's requirements on explanations), and since the idealized details are irrelevant, the derivation explains the proportionality between a pendulum's period and length .

4.3 Ineliminable Idealizations Revisited

With an account in hand of what is special about idealized explanation, along with a distinction between distorting and abstracting idealizations, it is possible to show how the ineliminably idealized explanations from Chapter Three are explanatory. This task divides into three components. The first task is to show that the interpretation of the thermodynamic limit and Boltzmann-Grad limits as abstractions, rather than as distortions, avoids the paradoxes of phase transitions and irreversibility, respectively. The second task is to show that, given this reinterpretation, the mathematical elements in the accounts of phase transitions and irreversibility remain well-defined. The third and final task is to show that, given the reinterpretation of the idealizing limits, it is plausible to construe the accounts of phase transitions and irreversibility as being explanatory despite being ineliminably idealized; this will be accomplished by appealing to the provisional (and partial) account of idealized explanation developed earlier in this chapter. These tasks are addressed first for the statistical mechanical account of phase transitions, and then for the Boltzmannian account of irreversibility.

4.3.1 Avoiding the Paradox of Phase Transitions

Recall the statistical mechanical explanation of the occurrence of phase transitions. Statistical mechanics identifies phase transitions as singularities in the Helmholtz free energy per particle of a system. It is a mathematical fact about the Helmholtz free energy per particle that it develops singularities only in the thermodynamic limit, which is the limit in which the system's particle number $N \to \infty$, the system volume $V \to \infty$, and the system's density N/V remains non-zero and finite. Any statistical mechanical account of why a system undergoes a phase transition must treat the system as if it exists in the thermodynamic limit. The necessity of appealing to this limit makes any statistical mechanical account of phase transitions an ineliminably idealized explanation.

The ineliminability of the thermodynamic limit raises the paradox of phase transitions. Every real system has only finitely many particles; and some of these systems undergo phase transitions. Yet, according to statistical mechanics, a system undergoes a phase transition only if it is in the thermodynamic limit. If systems in the thermodynamic limit are systems with infinitely many particles, then, no phase transitions occur in real systems.

It is possible to avoid the paradox of phase transitions by interpreting the thermodynamic limit as an abstraction rather than as a distortion, because such an interpretation does not entail that systems in the thermodynamic limit have infinitely many particles. Prior to interpreting this limit as an abstraction, it is important to note that the limit is typically taken *in the sense of van Hove.* Consider a sequence of bounded open regions within a region Λ in three-dimensional space. Let $V(r)$ be the volume of the set of points within a distance of r from the boundary of Λ, and let $V(\Lambda)$ be the volume of the region Λ. Then the limit $N(\Lambda) \rightarrow \infty$, $V(\Lambda) \rightarrow \infty$ is taken in the sense of van Hove just if, for every distance $r > 0$, the ratio $(V(r)/V(\Lambda)) \rightarrow 0$ as $N \rightarrow \infty$, where $N(\Lambda)$ is the number of particles in Λ. From now on, the thermodynamic limit as it occurs in the account of phase transitions will be understood in the sense of van Hove.

When the thermodynamic limit is treated as a distortion (and taken in the sense of van Hove), it is the limit in which the surface and boundary effects for the system

– the effects due to interactions that involve the walls or open surfaces of the system – vanish. This is because these effects occur near the boundaries of the system, and when the limits $N(\Lambda) \rightarrow \infty$ and $V(\Lambda) \rightarrow \infty$ are taken in the sense of van Hove and treated as distortions, the surface and boundary areas of the system come to occupy a vanishingly small proportion of the entire volume of the system. Since surface and boundary effects are not vanishingly small in real systems, real systems do not exist in the thermodynamic limit taken in the sense of van Hove.

When the thermodynamic limit is treated as an abstraction (and taken in the sense of van Hove), it is the limit in which the surface and boundary effects for the system are ignored or set aside; and a description of a system in this limit is a partial description of the system that sets aside details about the amount of surface and boundary effects in the system. But the limit does not ignore the finite and non-zero density of the system, nor the non-zero number of particles in the system, since the thermodynamic limit is neither a limit in which $N/V \rightarrow 0$ nor a limit in which $N \rightarrow 0$.

If the thermodynamic limit is an abstracting idealization, then real systems can exist in the thermodynamic limit: if the limit is an abstraction, to say that a system exists in the thermodynamic limit is to say that the system is described in a way that ignores the surface and boundary effects in the system. Since real systems can be described in this way, via partial descriptions, real systems can exist in the thermodynamic limit if that limit is an abstracting idealization. This is the same as saying that, as an abstracting idealization, the thermodynamic limit does not attribute properties to real systems that they do not have. (For example, as a distortion the thermodynamic limit attributes to systems the property of having a vanishingly small amount of surface effects; but as an abstraction, the limit merely ignores these effects.) Hence, the paradox of phase transitions is blocked when the thermodynamic limit is interpreted as an abstraction rather than as a distortion, because this interpretation does not make it the case that systems in the thermodynamic limit cannot be real systems – i.e., the interpretation falsifies the fourth premise in the paradox, according to which systems in the thermodynamic limit have infinitely many particles.

This result may be illustrated through a consideration of the Gibbs free energy per particle for systems in the thermodynamic limit. For a system of N particles, its Gibbs free energy per particle is defined to be the Gibbs free energy of the system, G, divided by the number of particles in the system. It is standard to represent the Gibbs free energy per particle for a system in the thermodynamic limit as a function of the pressure and temperature of the system: $g = g(p, T)$. The Gibbs free energy per particle for a system not in the thermodynamic limit contains more terms. Terrell Hill suggests that, for at least some systems not in the thermodynamic limit, the Gibbs free energy per particle is something like

$$g^* = g(p, T) + \frac{1}{\sqrt[3]{N}} a(p, T) + \frac{\ln N}{N} b(T) + \frac{1}{N} c(p, T),$$

where the term $a(p, T)$ describes the surface free energy and the last two terms describe effects due to rotational and translational motion. (These terms describe surface and boundary effects.)[72] There are no singularities in g^*. But in the thermodynamic limit, g^* reduces to g (all terms except the first vanish):

$$\lim_{N,V \to \infty, \frac{N}{V} = const} g^* = g = g(p, T).$$

[72]Hill suggests this as the Gibbs free energy for a single colloidal particle; see [46], pp. 1, 41-44.

The thermodynamic limit removes, from a full description of the Gibbs free energy per particle of a system, the terms that represent surface effects and effects due to rotational and translational motion.[73]

As an abstracting idealization, the thermodynamic limit transforms a full description of the Gibbs free energy per particle of a system into a description that ignores boundary and finite-size effects. The resultant Gibbs free energy per particle can develop singularities and thereby indicate the occurrence of phase transitions. But this Gibbs free energy per particle describes real systems, not systems with infinitely many particles; it is a description of real systems that is incomplete in virtue of ignoring the surface and boundary effects of real systems.

4.3.2 Mathematics in the Account of Phase Transitions

Recall that statistical mechanics identifies phase transitions as singularities in the Helmholtz free energy *per particle*. As the discussion of the Ising model in Chapter Three demonstrates, this function is well-defined in the thermodynamic limit. This is because the existence of the thermodynamic limit guarantees that the free energy per particle, or the free energy per site (of a lattice), is well-defined in the thermodynamic limit.

This is true even when the thermodynamic limit is interpreted as an abstraction. As an abstraction, the limit ignores the surface and boundary effects in the system. This is compatible with *not* ignoring the Helmholtz free energy per particle of the system. As an abstraction, application of the thermodynamic limit to the Helmholtz free energy per particle of a system yields a function for the Helmholtz free energy

[73] Joel Lebowitz makes a similar remark: "The advantage of this idealization [i.e., the thermodynamic limit] is that boundary and finite-size effects present in real systems, which are frequently irrelevant to the phenomena of interest, are eliminated in the thermodynamic limit" ([67], p. S351).

per particle of the system that ignores the surface and boundary effects in the system: the idealized function for the Helmholtz free energy per particle provides a partial description of the system. Importantly, this partial description is capable of containing singularities. The discussion of the Ising model in the previous chapter substantiates this claim: there are some temperatures at which there is a divergence of the free energy per site of an Ising system in the thermodynamic limit.

4.3.3 The Account of Phase Transitions as Explanatory

Having shown that the thermodynamic limit (taken in the sense of van Hove) can be interpreted as an abstracting idealization, that this interpretation avoids the paradox of phase transitions, and that various mathematical properties of the Helmholtz free energy per particle remain well-defined under this interpretation, it remains to show that it is plausible to construe the statistical mechanical account of phase transitions as being explanatory, despite that account's being ineliminably idealized.

According to the discussion of the previous section, there is a relevance requirement on idealized explanations, to the effect that any properties of a system that are idealized must be irrelevant to the phenomenon of interest if an idealized description of that system is to be explanatory of the phenomenon of interest. With respect to the statistical mechanical account of phase transitions, this requirement entails that the size and volume of systems in which phase transitions occur (or, at least, the values of these properties) must be irrelevant to the occurrence of those phase transitions. Since idealizing the size and volume of a system is tantamount to idealizing away surface effects on the system and effects due to translational and rotational motion (see the previous discussion of the Gibbs free energy per particle for systems in the thermodynamic limit), the relevance requirement is the requirement that these effects be irrelevant to the occurrence of phase transitions.

The relevance of these effects to the occurrence of phase transitions may be determined by an appeal to the preceding account of relevance. Consider a system that undergoes a phase transition. Step 1: Consider the nomic web in which this system is embedded. A veridical nomic model for the phase transition in this system is a representation of some portion of the nomic web in which the system is embedded that suffices to nomically produce the phase transition; this representation is given by the Helmholtz free energy per particle for the system (or, to say the same thing, by the partition function per particle for the system). Step 2: Even when details about the surface and boundary effects are removed from this model (e.g., by an abstracting idealization such as the thermodynamic limit taken in the sense of van Hove), the resultant residual model still manages to entail that the system undergoes a phase transition under certain conditions. Hence, boundary and surface effects are irrelevant to the occurrence of phase transitions, according to the preceding account of relevance. Given the plausibility of that account of relevance, it seems to be permissible to conclude that, at least sometimes, those properties of real systems that are idealized by the thermodynamic limit are irrelevant to the occurrence of phase transitions.[74]

In addition to satisfying the relevance requirement, the statistical mechanical account otherwise seems to be explanatory. The explanandum of the account is that

[74]One corollary of this way of thinking is that the consequence relation used in the account of relevance must be non-monotonic: if one adds premises about the details that are omitted by the ineliminable thermodynamic limit to the nomic model for the system in question, one fails to obtain the conclusion that the system undergoes a phase transition. The significance of this result remains to be explored.

a real system can undergo a phase transition. Generically, the statistical mechanical explanation of this explanandum proceeds according to the following steps:

1. The real system is represented by a microscopic model. Part of this model is an expression for the energy E of the system.

2. The partition function Q for the system is calculated, given a specification of the energy E.

3. The bridge law $\ln Q = -F/k_B T$ yields an expression for the Helmholtz free energy F of the system.

4. Application of the thermodynamic limit to the Helmholtz free energy yields an expression for the Helmholtz free energy per particle (or per site) of the system.

5. Expressions for various thermodynamic quantities, such as the specific heat of the system, are obtained by taking derivatives of the Helmholtz free energy per particle.

6. Mathematical analysis is used to show that there is a singularity in one of these thermodynamic quantities. These singularities are taken to correspond to the occurrence of a phase transition in the system.

This derivation appeals to the bridge law that relates the partition function of the system to the system's Helmholtz free energy, to the idealization of the thermodynamic limit, and to non-nomological conditions that pertain to the microscopic constitution of the system and the system's energy. And the derivation instantiates a common argument pattern within statistical mechanics. Consequently, it is plausible to suppose that the statistical mechanical account of phase transitions is explanatory, despite its being ineliminably idealized.

The explanation that the Ising model provides for the existence of ferromagnetic phase transitions in ferromagnets illustrates this general pattern. (For specific details, see Chapter Three.) The Ising model represents a ferromagnet as a lattice consisting of a fixed set of regularly-spaced sites connected to each other by bonds. The Ising model assumes an expression for the total energy of each configuration of this lattice, and this expression specifies the canonical partition function for the lattice. The partition function, along with a bridge law that relates the partition function per site to the Helmholtz free energy per site, yields an expression for the Helmholtz free energy per site of the lattice. In the thermodynamic limit, it can be shown that the Helmholtz free energy per site is not represented by an analytic function, and this non-analyticity corresponds to a logarithmic divergence in the specific heat of the ferromagnet. This divergence corresponds to a ferromagnetic phase transition in the magnet. Hence, from a bridge law that relates the partition function per site to the Helmholtz free energy per site, along with the idealization of the thermodynamic limit and non-nomological assumptions about the microscopic constitution of ferromagnets, the Ising model allows for a derivation of the explanandum statement, to the effect that there can be ferromagnetic phase transitions in ferromagnets.

4.3.4 Avoiding the Paradox of Irreversibility

Recall now the Boltzmannian explanation of irreversibility. The Boltzmannian account begins with a theorem of classical mechanics known as Liouville's theorem and derives Liouville's equation. Given certain assumptions about a system, concerning its interaction potentials and the like, one derives the BBGKY hierarchy from Liouville's equation. The BBGKY hierarchy yields the Boltzmann equation, under the statistical assumption that molecular chaos is present in the system initially, the Boltzmann-Grad limiting idealization, and other less important assumptions. By appealing to properties of the Boltzmann equation, it is possible to prove Boltzmann's H-theorem, which shows that there is an overwhelming likelihood for a non-equilibrium system irreversibly to approach an equilibrium state. The appeal to the Boltzmann-Grad limit is essential to this account of irreversibility. It renders the irreversible behavior predicted by the Boltzmann equation consistent with Poincaré's theorem. For this reason, the Boltzmannian account of irreversibility is an ineliminably idealized explanation.

The ineliminability of the Boltzmann-Grad limit raises the paradox of irreversibility. Every real system has only finitely many particles; and some of these systems exhibit irreversible behavior. Yet, according to the Boltzmannian account, a system exhibits irreversible behavior only if it exists in the Boltzmann-Grad limit. If systems in the Boltzmann-Grad limit have infinitely many particles, no real systems behave irreversibly.

It is possible to avoid the paradox of irreversibility by interpreting the Boltzmann-Grad limit as an abstraction rather than as a distortion, because such an interpretation does not entail that systems in the Boltzmann-Grad limit have infinitely many particles. Prior to interpreting this idealization as an abstraction, it will be useful to provide further information about what is accomplished, from a mathematical point of view, by taking the Boltzmann-Grad limit.[75]

[75] My discussion here would be improved if I were to find some discussion of the usual sense in which the Boltzmann-Grad limit is taken. The literature that appeals to the Boltzmann-Grad limit

Systems in the Boltzmann-Grad limit have the following mathematical properties:

- The particle number $N \to \infty$ while each particle's diameter $d \to 0$, in such a way that the quantity Nd^2 remains finite and non-zero.

- Each particle's mass $m \to 0$, but the system's total mass $M = Nm$ remains finite and non-zero.

- The volume of the container occupied by the system's particles $Nd^3 \to 0$ (since Nd^2 remains finite and non-zero but $d \to 0$).

- The average inter-particle distance $(V/N)^{\frac{1}{3}} \to 0$ (since $N \to \infty$ and V remains finite and non-zero).

- The system's volume V and mean free path (average distance particles travel between collisions, roughly equal to V/Nd^2) remains finite and non-zero (since both V and Nd^2 remain finite and non-zero).

- The number of collisions per unit time remains finite and non-zero (since Nd^2 remains finite and non-zero and, at least for a system of hard spheres, the number of collisions per unit time is given by $Nd^2\pi$.)

- For a finite interval of time, the total number of collisions remains finite and non-zero (since the number of collisions per unit time remains finite and non-zero).

- The probability of an arbitrarily selected particle undergoing a collision "Prob(single collision)" $\to 0$ (since, if it were not the case that "Prob(single collision)" $\to 0$,

in accounting for irreversibility tends to state that the limit is taken in some sense, without giving the details.

the total number of collisions over a finite period of time would not remain finite and non-zero).

If the Boltzmann-Grad limit is a distortion, then systems in the Boltzmann-Grad limit are systems that have infinitely many particles. Each particle of such a system has a vanishingly small mass and diameter. There are two important consequences of these distortions. The first is that the average distance between particles is vanishingly small; such a system is a sort of continuum, since a continuum is a system in which there is no distance between any of the system's components. The second consequence is that there is a vanishingly small probability for each particle to undergo a collision. These are all properties that real systems do not have: in real systems, there is a finite inter-particle distance and a finite probability for each particle to undergo a collision. If the Boltzmann-Grad limit is a distortion, systems in the Boltzmann-Grad limit are not real systems.

If the Boltzmann-Grad limit is an abstraction rather than a distortion, then a description of a system in the Boltzmann-Grad limit is a partial description of the system. This partial description does not attribute to the system a finite inter-particle distance and a finite probability for each particle to undergo a collision (or, at least, does not attribute to the system values for these properties). Taking the Boltzmann-Grad limit has the effect of ignoring these properties of real systems (or, at least, ignoring the values of these properties). And as an abstraction rather than a distortion, taking the Boltzmann-Grad limit does *not* have the effect of attributing to systems properties that real systems do not have, because ignoring finite inter-particle distances and single particle collision probabilities does not entail claiming that such distances and probabilities are arbitrarily close to zero.

If the Boltzmann-Grad limit is an abstracting idealization, then real systems can be in the Boltzmann-Grad limit: to say that a system is in the Boltzmann-Grad limit is to say that the system is described in a way that ignores the average inter-particle distance and the probability of a particle undergoing a collision. Since real systems can be described in this way, via partial descriptions, real systems can be in the Boltzmann-Grad limit if that limit is an abstracting idealization. This is all the same as saying that, as an abstracting idealization, the Boltzmann-Grad limit does not attribute properties to real systems that real systems do not have. Hence, the paradox of irreversibility is blocked by treating the Boltzmann-Grad limit as an abstraction rather than a distortion, because this interpretation of the limiting idealization does not make it the case that systems in the Boltzmann-Grad limit have properties that real systems do not have – i.e., the interpretation falsifies the fourth premise in the paradox, according to which systems in the Boltzmann-Grad limit have infinitely many particles.

This result may be illustrated through a comparison of the equations of the BBGKY hierarchy to the equations of the Boltzmann hierarchy (obtained by applying the Boltzmann-Grad limit to each equation of the BBGKY hierarchy).[76] I introduce the following symbols. Let $\overrightarrow{\xi_i}$ represent the velocity of the i-th particle prior to its collision with some other particle and $\overrightarrow{V_{ij}}$ represent the velocity of the i-th particle relative to an arbitrary j-th particle ($i \neq j$). Let f'_{s+1} be the function obtained from f_{s+1} by replacing $\overrightarrow{\xi_i}$ and $\overrightarrow{\xi_j}$ in the latter (velocities after collision) with $\overrightarrow{\xi_i}'$ and $\overrightarrow{\xi_j}'$ in the former (velocities before collision), respectively. Let q_i represent

[76]These equations are given by [20], pp. 49-53.

the center of the i-th particle and q_j the center of an arbitrary j-th particle; $\overrightarrow{n_{ij}}$ is the unit vector directed along the line that joins the centers of the i-th and j-th particles.

For an N-particle system, the generic equation for each member of the BBGKY hierarchy is:

$$\frac{\partial f_s}{\partial t} + \sum_{i=1}^{s} \overrightarrow{\xi_i} \cdot \frac{\partial f_s}{\overrightarrow{q_i}} = (N - s)\sigma^2 \sum_{i=1}^{s} \int [f'_{s+1} - f_{s+1}] |\overrightarrow{V_{ij}} \cdot \overrightarrow{n_{ij}}| d\overrightarrow{n_{ij}} d\overrightarrow{\xi_j},$$

for $s = 1, \ldots N$. The term $\frac{\partial f_s}{\partial t}$ gives the time variation of f_s; the drift term $\overrightarrow{\xi_i} \cdot \frac{\partial f_s}{\overrightarrow{q_i}}$ gives the spatial variation of f_s for particle i. The term on the right hand side, known as the collision term, gives the effect of collisions between particles. This collision term is a sum of the effect of collisions between two particles, the effect of collisions between three particles, and so on, up to the effect of collisions between all N particles. Note that the collision term for the s-particle partial probability distribution function of the BBGKY hierarchy depends upon the $s + 1$-particle partial probability distribution function. This is why the equations of the BBGKY hierarchy are open equations (except for the case where $s = N$).

The generic equation for each member of the Boltzmann hierarchy is obtained from the generic equation for each member of the BBGKY hierarchy by applying the Boltzmann-Grad limit to each of the latter equations. To do this, we transform the right-hand side of the previous equation so that it has two terms and is equal to:

$$N\sigma^2 \sum_{i=1}^{s} \int [f'_{s+1} - f_{s+1}] |\overrightarrow{V_{ij}} \cdot \overrightarrow{n_{ij}}| d\overrightarrow{n_{ij}} d\overrightarrow{\xi_j}$$

$$-s\sigma^2 \sum_{i=1}^{s} \int [f'_{s+1} - f_{s+1}] |\overrightarrow{V_{ij}} \cdot \overrightarrow{n_{ij}}| d\overrightarrow{n_{ij}} d\overrightarrow{\xi_j}.$$

Since the quantity $s\sigma^2 \to 0$ as $d \to 0$ but the quantity $N\sigma^2$ remains finite and non-zero in the Boltzmann-Grad limit, the second term of this equation vanishes in the

Boltzmann-Grad limit. Hence, the result of applying the Boltzmann-Grad limit to each equation in the BBGKY hierarchy is:

$$\frac{\partial f_s}{\partial t} + \sum_{i=1}^{s} \overrightarrow{\xi_i} \cdot \frac{\partial f_s}{\overrightarrow{q_i}} = (N\sigma^2) \sum_{i=1}^{s} \int [f'_{s+1} - f_{s+1}] |\overrightarrow{V_{ij}} \cdot \overrightarrow{n_{ij}}| d\overrightarrow{n_{ij}} d\overrightarrow{\xi_j},$$

for $s = 1, \ldots N$.

As an abstracting idealization, the Boltzmann-Grad limit removes some of the details about the effect of collisions between particles. This may be observed by comparing the right-hand sides of the generic equations for members of the BBGKY hierarchy and the Boltzmann hierarchy, respectively. The Boltzmann equation itself, which characterizes irreversible behavior, is subsequently obtained from the equations of the Boltzmann hierarchy by invoking the assumption of initial molecular chaos. A secondary role of the Boltzmann-Grad limit is to ensure that this initial molecular chaos propagates; that is, the Boltzmann-Grad limit also functions to ignore details about correlations that come to exist between particles when they collide.

4.3.5 Mathematics in the Account of Irreversibility

The Boltzmann equation remains well-defined in the Boltzmann-Grad limit under the interpretation that treats that limit as an abstraction. Recall that the Boltzmann equation describes the time evolution of a 1-particle partial probability density function f_1. This function gives the probability density of finding one randomly chosen particle, of an N particle system, in some subregion of μ-space, without regard to the locations of the other $N-1$ particles in that phase space. Specifically, where x_1 represents, in μ-space, the generalized coordinates and conjugate momentum of a particle randomly chosen from an N-particle system, $f_1(x_1, t)dx_1$ gives the probability density of finding that particle in the volume of μ-space between x_1 and $x_1 + dx_1$ at time t. An important fact about the 1-particle partial probability density function of an N-particle system, yet to be noted, is that it is equal to a function f^* divided by the total mass M of the system:

$$f_1 = \frac{f^*}{M}.$$

The function f^* is the expected mass density in μ-space, the expected mass per unit volume in Γ-space (see [20], p. 57). There is a normalization condition on f^*, to the effect that its integral over μ-space be equal to the total mass of the system. The total mass of the system, of course, is equal to the sum of the masses of each particle in the system.

The only way for the total mass to remain finite in the $N \to \infty$ limit is to take the limit in which each particle's mass $m \to 0$, requiring that, in this combined limit $N \to \infty$ and $m \to 0$, M remains a finite and non-zero constant (see [20], p. 43). This requirement can be met for the interpretation of the idealizing limit $m \to 0$ as an abstraction, by noting that it is possible to ignore a feature of the parts of a system without ignoring a feature that the system has as a whole. For instance, it is possible to ignore the amount of weight of each marble in a bag of marbles without ignoring the amount of weight of the entire bag of marbles. Similarly, it is possible to ignore the amount of mass of each particle of a system without ignoring the amount of total mass of the system.

Since M remains finite in the Boltzmann-Grad limit, the function f^* is well-defined in that limit. Since the 1-particle partial probability density function is equal to the expected mass density f^* divided by the total mass of the system, the 1-particle partial probability density function also remains well-defined in the Boltzmann-Grad limit.

And since the Boltzmann equation describes the time evolution of the 1-particle partial probability density function, the Boltzmann equation itself remains well-defined in the Boltzmann-Grad limit. These claims are true even if the Boltzmann-Grad limit is an abstraction.

4.3.6 The Account of Irreversibility as Explanatory

If the Boltzmann-Grad limit is an abstracting idealization, then the Boltzmann equation is a partial description of real systems, a description obtained (in effect) by ignoring some of the effects of collisions between particles. This incomplete description happens to describe irreversible behavior. The effect of the collisions not described by the equations of the Boltzmann hierarchy, and subsequently not represented by the Boltzmann equation, is given as

$$-(s\sigma^2)\sum_{i=1}^{s}\int [f'_{s+1} - f_{s+1}]|\overrightarrow{V_{ij}} \cdot \overrightarrow{n_{ij}}| d\overrightarrow{n_{ij}} d\overrightarrow{\xi_j}.$$

Having shown that the Boltzmann-Grad limit can be interpreted as an abstraction, that this interpretation avoids the paradox of irreversibility, and that the Boltzmann equation remains well-defined under this interpretation of the Boltzmann-Grad limit, it remains to show that it is plausible to construe the Boltzmannian account of irreversibility as explanatory, despite that account's being ineliminably idealized.

With respect to the Boltzmannian account of irreversibility, the relevance requirement on idealized explanations entails that the size and particle diameter of systems that exhibit irreversibility (or, at least, the values of these properties) must be irrelevant to the occurrence of irreversible behavior. Since idealizing the size and particle diameter of a system is tantamount to idealizing away some of the effects of collisions between particles, the relevance requirement entails that these effects are irrelevant to the occurrence of irreversible behavior.[77] The relevance of these effects to irreversibility may be determined by an appeal to the preceding account of relevance. Consider a system that exhibits irreversible behavior. Step 1: Consider the nomic web in which this system is embedded. A veridical nomic model for the irreversible behavior in this system is a representation of some portion of the nomic web in which the system is embedded that suffices to nomically produce the irreversible behavior; this representation is given by the equations of the BBGKY hierarchy. Step 2: Even when details about the effects of collisions between particles are removed from this model (e.g., by an abstracting idealization such as the Boltzmann-Grad limit), the resultant model (represented by the Boltzmann equation) entails that the system exhibits irreversible behavior. Hence, these effects due to inter-particle collisions are irrelevant to the occurrence of irreversibility, according to the preceding account of relevance. Given the plausibility of that account, it seems to be permissible to conclude that, at least sometimes, those properties of real systems that are idealized by the Boltzmann-Grad limit are irrelevant to the occurrence of irreversible behavior.

In addition to satisfying the relevance requirement, the statistical mechanical account otherwise seems to be explanatory. The explanandum is that a real system can exhibit irreversible behavior. Generically, the Boltzmannian account of this explanandum proceeds according to the following steps:

1. Hamilton's equations of motion can be used to obtain the Liouville equation for the system of interest, which governs the time evolution of the probability density function for the system.

[77] Situations in which these effects are irrelevant include the behavior of a moderately rarefied gas such as the upper atmosphere; this behavior is important in planning space shuttle re-entries.

2. Under certain assumptions (about interaction potentials, etc), the Liouville equation entails a set of N integro-differential equations; these equations form the BBGKY hierarchy.

3. The Boltzmann-Grad limit is applied to each equation of the BBGKY hierarchy, yielding a set of equations known as the Boltzmann hierarchy.

4. The Boltzmann equation is obtained from the Boltzmann hierarchy by appealing to the assumption of initial molecular chaos and the result that initial molecular chaos propagates in the Boltzmann-Grad limit.

5. The H-theorem is obtained from analysis of the Boltzmann equation. This theorem shows that the entropy of virtually every system increases monotonically in time until the system obtains a state of equilibrium.

The Boltzmann equation is obtained from the law-like equations of classical mechanics as well as various modelling assumptions and idealizations, and the H-theorem is a consequence of the Boltzmann equation. The derivation of this equation instantiates a common argument pattern within statistical mechanics. And the equation itself traces the causal interactions that produce irreversible behavior.

4.4 Conclusion

James Woodward suggests five desiderata for any satisfactory philosophical account of explanation ([123], p. 23):[78]

1. Descriptive adequacy: The account should capture relevant features of paradigmatic instances.

[78]These desiderata are similar to those given by Michael Friedman ([35]), Philip Kitcher ([56], p. 508), and R.I.G. Hughes ([47]).

2. Philosophical Adequacy: The account should provide insight into the features of explanations that allow them to convey information.

3. Dialectical Superiority: The account should solve problems not adequately dealt with by previous accounts.

4. Epistemic Accessibility: The account should give a plausible story about how people who use such explanations can learn information given by them.

5. Normative Guidance: The account should allow for an evaluation of different explanations.

The partial account of idealized explanation provided in this chapter, which treats idealizations as abstractions, satisfies the first three of these desiderata, at least with regard to the conditions on explanation that pertain specifically to idealizations. (Showing how the account satisfies the other two desiderata is a project for some other time.)

The account appears to be descriptively adequate. It accommodates idealized explanations involving the simple pendulum, a paradigmatic idealized system. The account can also claim success with respect to the statistical mechanical explanation of phase transitions and the Boltzmannian explanation of irreversibility. These cases not only further support the claim that the account is descriptively adequate, but also render the account dialectically superior to other philosophical accounts of idealized explanation, because other accounts cannot accommodate ineliminably idealized explanations (see Chapter Three for details). Finally, the account is philosophically adequate, because it shows that idealized explanations convey explanatory information in virtue of being partial descriptions that satisfy various explanatory conditions. Idealized explanations explain in the same way that incomplete explanations explain.

The descriptive and philosophical adequacy of the above account of idealized explanation, and its dialectical superiority over rival philosophical accounts of idealized explanation, are strong evidence in favor of the account. Since interpreting idealizations as abstractions forms the core of this new account of idealized explanation, the evidence in favor of the account is also evidence in favor of interpreting idealizations as abstractions. And since this account succeeds where the accounts surveyed in Chapter Two fail, such an account of idealized explanation is better than those from Chapter Two.

The moral of Chapter One was that any successful account of idealized explanation must either abandon the interpretation of idealizations as distortions or allow some falsehoods to be explanatory. Chapter Two outlined accounts of idealized explanation that retain the interpretation of idealizations as distortions. Chapter Three showed that such an interpretation is incompatible with the existence of ineliminably idealized explanations. This chapter provides an account of idealized explanation that abandons the interpretation of idealizations as distortions, in favor of an interpretation of idealizations as abstractions. On this account, idealized explanations are a species of incomplete explanation, in which the incompleteness is, in some sense, simplifying. Such an account accommodates paradigmatic cases of idealized explanation, as well as cases of ineliminably idealized explanation. This favors an interpretation of ineliminably idealizations as abstractions rather than distortions, and undermines the necessity of showing how idealized descriptions can be explanatory despite being false.

CHAPTER 5

ERROR THEORIES, EARMAN'S PRINCIPLE, AND EMERGENCE

Sometimes only partial (incomplete) descriptions correctly describe systems. This is a consequence of the presumption that there are ineliminably idealized explanations and the interpretation of idealizations as abstractions. If P is the explanandum of an ineliminably idealized explanation, then P must occur in some real system, given the presumption that every explanandum obtains (every explanandum is true). Let one of the real systems in which P occurs be system S. Then, since P is the explanandum of some ineliminably idealized explanation, S must exist in some idealizing limit. Hence, some real systems must be systems in idealizing limits – that is, sometimes the correct description of a system requires the omission of details about the system. To put this as a slogan: completeness does not guarantee correctness.

This corollary conflicts with what Robert Batterman dubs the "traditional" view about models in the physical and applied mathematical sciences, according to which "a model is better the more details of the real phenomenon it is actually able to represent mathematically" ([5], p. 21). Batterman argues that this traditional view is mistaken, because sometimes the explanation of why different systems exhibit the same pattern requires an appeal to minimal models, models that do not represent every detail of the systems of interest. The preceding argument concurs with Batterman's conclusion, holding that the traditional view is mistaken because some explanations are ineliminably idealized, and the explananda of such explanations require descriptions (or models) that are incomplete. Moreover, the preceding argument extends Batterman's objection to the traditional view. Batterman's argument concerns multiply realized patterns such as the fact that struts buckle upon reaching a critical load. The argument from Chapters Three and Four, however, concerns individual events (or classes of events) such as the occurrence of phase transitions in individual systems. So the traditional view is mistaken not only for multiply realized patterns, but also for some individual (classes of) events.

This result depends upon the presupposition that there are ineliminably idealized explanations, a presupposition which is by no means uncontroversial; in fact, there are several arguments that purport to establish its falsity. One aim of this chapter is to show these arguments to be ineffectual. A further aim is to show that the reasons invoked by the arguments against the presumption of ineliminably idealized explanations are less plausible than the reasons invoked in its favor, thereby legitimizing the appeal to the presumption on grounds of reflective equilibrium.

5.1 Error-Theoretic Objections

5.1.1 Two Arguments

Statistical mechanics represents the phase transitions of a system by mathematical singularities in the partition function for that system, and it accounts for the occurrence of a phase transition in a system by showing that the partition function for that system contains a singularity.[79] Roughly, the explanandum in this case is 'System S undergoes a phase transition', and statistical mechanics takes this to mean the same thing as 'The partition function for S develops a singularity'. The Boltzmannian account of irreversibility characterizes the tendency of non-equilibrium states of a system to approach equilibrium states as a tendency for the entropy of that system to increase irreversibly, and it accounts for this irreversibility by showing that the H-theorem holds of the system. Roughly, the explanandum in this case is 'The non-equilibrium states of system S have a tendency to irreversibly approach an equilibrium state', and the Boltzmannian approach takes this to mean the same thing as 'When S is in a non-equilibrium state, its H-function decreases monotonically (i.e., its entropy increases monotonically)'. In both cases, the systems of interest are real systems.

If these accounts of phase transitions and irreversibility are explanatory, then the explanandum of each account must be true of real systems; for it is a requirement on every explanation that its explanandum be true. Craig Callender provides arguments that purport to show these explananda to be false of real systems.

Callender argues that the statement 'The partition function for system S develops a singularity' cannot be true of any real system. For singularities in the partition

[79]This is not exactly correct. Statistical mechanics represents phase transitions as singularities in the partition function *per particle*, as shown in Chapter Three. But philosophers who write about phase transitions tend to overlook this detail and speak as if the partition function, rather than the partition function *per particle*, is the function relevant for identifying phase transitions. Instead of correcting this way of speaking through the chapter, I acquiesce in it, leaving the reader the responsibility of understanding that the "partition function" being discussed is, in fact, the partition function *per particle*.

function of any system can only occur in the thermodynamic limit, yet no real system exists in the thermodynamic limit because real systems only have finitely many particles ([16], pp. 548-550; see also [73], [125]). Moreover, this statement cannot be true of any real system, because the partition function for any system contains singularities only if there are no fluctuations in the system; yet real systems have fluctuations ([16], p. 550). This argument purports to show that partition functions for real systems cannot contain singularities. Since statistical mechanics represents phase transitions as singularities in partition functions, this argument further seems to show that, according to statistical mechanics, there are no phase transitions.

Callender further argues that the statement 'The entropy of a system increases monotonically unless that system is in equilibrium' cannot be true of real systems. For the entropy of any real system is a time-reversal invariant function of the dynamical variables of that system. Since any closed real system has a bounded phase space, the time-reversal invariance of the entropy function entails that a system's entropy cannot monotonically increase: "In short, if [entropy] is a function of the dynamical variables of an individual system, then [entropy] cannot exhibit monotonic behaviour" ([16], p. 543). Since the Boltzmannian account characterizes the tendency of non-equilibrium states of a system to approach equilibrium as a tendency for the entropy of the system to increase monotonically, this argument seems to show that, according to the Boltzmannian account, real systems in non-equilibrium lack a tendency to approach equilibrium.

5.1.2 Callender's Eliminativism

Callender does not take these arguments to show that statistical mechanics does not account for phase transitions or irreversibility. Rather, he takes the arguments to show that statistical mechanics mischaracterizes the explanandum. Rather than representing phase transitions as singularities in partition functions, Callender suggests that statistical mechanics should represent them as non-singular solutions that, in some sense, approximate singularities. He writes,

> Presumably, there are non-singular solutions to the partition function describing real systems that give rise to the macroscopic transitions called phase transitions. ... Analytic partition functions must govern the phase transition and in some sense approximate a singularity" ([16], p. 550).

This method of representing phase transitions does not require that systems be in the thermodynamic limit in order to undergo phase transitions.

Callender's position is not the isolated remark of a philosopher; physicists espouse it as well. Consider the extended remarks of the physicist J.E. Mayer:

> There is probably no other inclusive field of science in which it is more tempting to expect complete mathematical rigor from beginning to end than in equilibrium statistical mechanics. The axioms are the laws of mechanics, which, for molecular systems at least, can be put in concise mathematical form. The end product is an equally concise set of a very few (2 or 3) mathematically formulated laws, those of thermodynamics. It appears to this author, however, that a search for complete rigor in the usual mathematical form is illusory, and, when pursued too industriously, has more often led to obscurantism than to clarity.
>
> The first difficulty that arises is that the conciseness and precision of mathematical formulation of the laws of thermodynamics are actually invalid for real systems of finite size; average values are not identical with the most probable, the probability of a spontaneous measurable decrease in entropy is infinitesimal but not zero, and *phase changes are not singularities but merely finite changes in derivatives within an inobservably infinitesimal, but non-zero, range of variables.* The demand that one treats only infinite

> systems obviates these difficulties, but leaves precious little applicability of thermodynamics to the real world ([78], pp. 236-237, emphasis added).

Like Callender, Mayer suggests that the representation of phase transitions as singularities is an artifact of an over-reliance on mathematical rigor. A proper explanation of phase transitions requires an account that is less precise and thereby able to avoid an appeal to the thermodynamic limit.

Callender adopts a similar strategy as regards irreversibility. Rather than characterize the entropy of non-equilibrium systems as a monotonically increasing function, statistical mechanics should hold that the entropy of non-equilibrium systems does not decrease for very long observational time scales ([16], p. 544). This re-characterization of the behavior of real systems in non-equilibrium is compatible with their time-reversal invariance and the Poincaré recurrence theorem, because a system's entropy not decreasing for very long observational time scales is consistent with the system's entropy decreasing over an even longer cosmic time scale. This consistency is bought at the price of precision, however, since what counts as "very long" is somewhat vague.[80]

The physicist Paul Davies espouses a position similar to Callender's. According to Davies,

> there can be no true 'equilibrium' state for [an isolated box of gas]. The state that we have called equilibrium is only the most frequented state, and does not satisfy the usual criterion of equilibrium because the system will leave it eventually (though only appreciably after vast periods of time). For this reason, all statements about equilibrium and irreversibility

[80]Nonetheless, Callender takes the re-characterization to be preferable to abandoning the Boltzmannian approach in favor of a Gibbsian approach, because even if the Gibbsian approach shows that some function of an ensemble of systems changes monotonically over time, such a function does not explain the thermal behavior of individual real systems because it is not a function of the dynamical variables of individual real systems ([16], p. 544).

> . . . should be interpreted as meaning on time scales much less than the Poincaré recurrence times ([28], pp. 59-60).

Like Callender, Davies suggests that a proper characterization of the behavior of non-equilibrium systems must allow for their entropy occasionally to decrease.

Callender's response to the arguments for the falsity of the explananda of the statistical mechanical account of phase transitions and the Boltzmannian account of irreversibility is to reject both the statistical mechanical analysis of what it is for a system to undergo a phase transition and the Boltzmannian characterization of the behavior of non-equilibrium systems. In their place, Callender advocates the substitution of a less specific, and thereby less mathematically precise, analysis of what it is for a system to undergo a phase transition, as well as a less specific, less precise characterization of the behavior of non-equilibrium systems, respectively.

5.1.3 Liu's Instrumentalism

Not everyone who accepts the cogency of Callender's arguments agrees with his response.[81] For instance, Chuang Liu agrees with Callender that the explanandum of the statistical mechanical account of phase transitions is false, and for essentially the same reasons. Liu further agrees that a system undergoes a phase transition when the partition function for the system approximates a singularity:

> . . . the singularities which represent [phase transitions] . . . should be regarded as mere artifacts or fictions of [the thermodynamic limit] that do not exist in reality. [Phase transitions] are just places where rather dramatic changes of thermo-variables take place so that their derivatives make

[81]According to Sang Wook Yi, the argument shows that "[thermodynamics] 'corrects' [statistical mechanics] so that [statistical mechanics] can accommodate certain experimentally verified phenomena" ([125], p. 1031). I am inclined to agree with this claim; one aim of this dissertation has been to show how statistical mechanics can accommodate such phenomena despite giving ineliminably idealized accounts of them.

> sharp, but not discontinuous, changes ([73], p. S336; brackets replace Liu's abbreviations with appropriate phrases).

However, Liu advocates the retention of the statistical mechanical analysis of what it is for a system to undergo a phase transition, due to the theoretical utility of the analysis. Rather than replace the analysis with a less precise one, Liu suggests a relaxation of the criteria for applying the analysis to real systems. He writes,

> Predicates we use to describe [phase transitions] and [critical points] in [the thermodynamic limit] of infinite systems in [statistical mechanics] are mathematical ones that result from an accentuation or exaggeration of the corresponding physical properties by neglecting or filling out negligible differences. With such predicates, scientists must demand strict exactness among their relations .... But when such predicates are applied to actual physical systems, estimates of approximation are brought back in so that the right kind of systems are picked out by the predicates. For instance, being a critical point as a mathematical predicate is defined by a singular point on an isotherm; but when using it to pick out a physical critical point a certain range of approximation to the singular point should be understood so that it picks out the right set of systems. This is similar to our use of most exact magnitudes, such as 'weighing 100 kg'. We are justified to use it to pick out objects whose weight is not exactly 100 kg but very close to it ([73], p. S340).

Liu is not alone in his position. R.A. Minlos, a mathematical physicist, expresses a similar attitude:

> ...several important notions of statistical physics can be rigorously defined only in the framework of the thermodynamic limit (for example, the important notion of phase transition). Of course, one has to remember that real physical systems are finite and the thermodynamic limit means some idealized discription [sic] of reality, but that can be said about any mathematical (or theoretical) method in physics ([82], p. 22).

Minlos and Liu agree that the representation of phase transitions as singularities is an idealization of reality. Their solution to this problem is not to revise the theoretical criteria for the representation of phase transitions, as Callender and Mayer advise. This is because such a representation is an inevitable result of representing the world with mathematics. Instead, allowance should be made for representing a complicated world with mathematical exactitude. This is an allowance made in the application of statistical mechanics to real systems, not an allowance made within the representational schemes of statistical mechanics.

Although Liu only applies his view to the statistical mechanical account of phase transitions, his position can be extended to the Boltzmannian account of irreversibility. According to such a position, the entropy of real systems does not increase monotonically in time. Nonetheless, this characterization of the approach to equilibrium, as a monotonic process, should be retained owing to its theoretical utility. Moreover, the criteria for applying this characterization to real systems should be relaxed, to allow for the inevitable discrepancies that arise from characterizing a complicated world with mathematical exactitude.

5.1.4 Callender and Liu as Error Theorists

Despite their differences, both Callender and Liu essentially advocate an error-theory about the predicates 'undergoes a phase transition' and 'exhibits irreversibility'.[82] According to Paul Boghossian, an error theory about a fragment of a discourse is a theory according to which, although the predicates of that fragment denote properties and thereby equip declarative sentences that contain those predicates with truth conditions, nothing actually exemplifies the properties denoted by those predicates, so that declarative sentences in the fragment of discourse are systematically false ([11],

[82] Callender and Liu's thesis may also be categorized as a version of what Kevin Davey calls the liberal view of mathematical rigor in physics, "according to which it is inappropriate to demand that even the most mature areas of physics traffic exclusively in mathematically rigorous concepts and arguments" ([27], p. 441).

p. 159). The discourse of statistical mechanics takes the predicate ‘undergoes a phase transition’ to be best explicated as ‘develops a singularity in the partition function’. (This interpretation of the predicate ‘undergoes a phase transition’ is neutral as to whether real systems instantiate the predicate.) For instance, according to statistical mechanics, saying that a system undergoes a phase transition is the same as saying that the system develops a singularity in its partition function. Furthermore, the discourse of statistical mechanics takes the predicate ‘exhibits irreversibility’ to be best explicated as ‘does not return to any previous state’. For instance, according to statistical mechanics (or, at least, the Boltzmannian approach), saying that the behavior of a system is irreversible is the same as saying that the system never returns to any of its previous states. An error theory about the declarative sentences of statistical mechanics that contain either the predicate ‘undergoes a phase transition’ or the predicate ‘exhibits irreversibility’ is, accordingly, a theory according to which those predicates denote properties that happen never to be exemplified by real systems – it is a theory according to which declarative sentences containing those predicates are systematically false.

Both Callender and Liu are error theorists about this fragment of statistical mechanics. Both maintain that, for any real system S, the sentence ‘S undergoes a phase transition’ is false; and both maintain that, for any real system S, the sentence ‘S exhibits irreversible behavior’ is false. Callender and Liu are, however, different kinds of error theorists. Liu is what Boghossian calls an instrumentalist. An instrumentalist grants that sentences containing certain predicates are systematically false, but maintains that the continued use of those sentences “serves an instrumental purpose that will not easily be discharged in some other way” ([11], p. 159). This is precisely the position Liu takes with respect to the predicate ‘undergoes a phase transition’: although its extension is empty, its theoretical utility counsels its retention.

Callender, in contrast, is what Boghossian calls an eliminativist. An eliminativist takes the systematic falsity of sentences involving certain predicates to be sufficient grounds for eliminating and replacing those predicates. This is precisely the position Callender takes with respect to the predicates ‘undergoes a phase transition’ and ‘exhibits irreversibility’. In effect, Callender’s proposal is that the former predicate be replaced with a new (albeit homophonic) predicate ‘undergoes a phase transition*’, where ‘undergoes a phase transition*’ is to be explicated as ‘develops an approximation to a singularity in the partition function’. Likewise, Callender’s proposal is that ‘exhibits irreversibility’ be replaced with a new (albeit homophonic) predicate ‘exhibits irreversibility*’, where ‘exhibits irreversibility*’ is to be explicated as ‘does not return to any of its previous states for very long observational time scales’.

If either Callender’s eliminativism or Liu’s instrumentalism is correct, then the statistical mechanical account of phase transitions and the Boltzmannian account of irreversibility are not explanatory. For, if either of these error-theories is correct, the putative explanandum of each account is false, thereby disqualifying from being explanatory any account that purports to show why the explanandum is true.[83] Moreover, if those accounts are not explanatory, there is no evidence that some explanations are ineliminably idealized (at least, no evidence provided within this dissertation). This would undermine the argument in favor of interpreting idealizations as abstractions (see Chapter Four), as well as the criticisms of the philosophical accounts

[83]Perhaps this result can be avoided, if there is an error-theoretic account of scientific explanation that does not require an explanandum to be true in order to be explained. But I am unaware of any such account.

of idealized explanation that allow idealizations to be explanatory despite being false (see Chapters Two and Three).

5.2 Responses to Error-Theories

There are two ways to respond to an error-theory. A direct approach contests one of the premises of the argument that establishes the error theory; an indirect approach contests the internal coherence of the error-theory itself (see [81], p. 98). Within the philosophical literature, Robert Batterman provides an indirect response to an error-theory about the predicate 'undergoes a phase transition'. The aim of this section is to assess the adequacy of Batterman's response, and to develop a direct response to the error-theories of Callender and Liu. The assessment of Batterman's response is that, if cogent, it establishes an incoherence in Callender's eliminativism but not in Liu's instrumentalism. (This is a limitation, but not a criticism, of Batterman's argument, since the argument is only directed at Callender's position on phase transitions to begin with.) The direct response to be developed, as a supplement to Batterman's indirect response about phase transitions, is that the arguments in favor of either kind of error-theory succeed only if idealizations are distortions – and fail if idealizations are abstractions. This direct response undercuts the motivation for both Callender's eliminativism and Liu's instrumentalism, and applies to error-theories about the predicates 'undergoes a phase transition' and 'exhibits irreversibility'.

5.2.1 Batterman's Indirect Response

In responding to Callender's eliminativism about the predicate 'undergoes a phase transition', Batterman distinguishes two kinds of discontinuities, mathematical and physical. Mathematical discontinuities are singularities in mathematical equations. For example, a singularity in the partition function is a mathematical discontinuity. Physical discontinuities in a system are qualitative changes in the system. For example, a drop of water breaking off from a stream of water is a physical discontinuity. According to Batterman, phase transitions are also physical discontinuities. For the phases of a system are differentiated by their qualitative features, and this strongly suggests that the different phases of a system are, in fact, qualitatively distinct from each other. This is corroborated by observations and by typical definitions of a phase. For instance, according to Bimalendu Roy, a phase is "defined as any homogeneous and physically distinct part of a system which is separated from other parts of the system by a definite boundary" ([98], p. 421). This definition suggests that phases are qualitatively distinct from each other, because it requires definite boundaries between different phases. Moreover, solid phases of systems sometimes have a lattice structure that is absent in non-solid phases of systems, and systems in a solid phase tend to occupy different volumes than the same systems in a liquid or gaseous phase.

From the claim that phase transitions are physical discontinuities, Batterman infers that phase transitions should be represented as mathematical discontinuities. The inference is licensed by Batterman's thesis about representation, to the effect that all physical discontinuities are to be represented as mathematical discontinuities. This thesis finds corroboration with the case of breaking streams of water, which are represented as mathematical discontinuities within the Navier-Stokes equations (see [8]).

Liu and Emch provide further reasons in favor of this thesis. They write,

> The difference between two phases, e.g., solid and liquid, is better captured by a singularity. Even before Gibbs, Maxwell had complemented the van

> der Waals model of real fluids by a construction, known as the 'Maxwell plateau,' the effect of which is to make the isotherms only piecewise analytic [i.e., discontinuous]. Should we regard this idealization as no more than an ad hoc construction? The construction was not only supported later by entirely rigorous mathematical arguments, it was also, first by Andrews, observed by successively cleaner and more precise experiments that the isotherms are better described as piece-wise analytic [i.e., mathematically discontinuous] rather than as having arbitrarily sharp corners" ([74], p. 155).

Liu and Emch seem to suggest that if one were not to represent physical discontinuities as mathematical discontinuities, one's system of mathematical representation would fail to distinguish, in the most perspicuous way possible, phenomena that are physically discontinuous from those that are continuous.

Batterman's argument may be summarized as follows. Since phase transitions are physical discontinuities, and since all physical discontinuities should be represented as mathematical discontinuities, phase transitions should be represented as mathematical discontinuities. Specifically, phase transitions should be represented as singularities in partition functions. Hence, Callender's suggestion, that phase transitions be represented as *approximations* to singularities in partition functions, is mistaken: it fails to represent phase transitions in the most perspicuous way possible, namely, as mathematical discontinuities.

Although Batterman's conclusion, if correct, shows that Callender's eliminativism is mistaken, it does not show that Liu's instrumentalism is mistaken. (Nor is it intended to.) For Liu's instrumentalism agrees that phase transitions should be represented as singularities in partition functions, on the grounds that this method of representation is theoretically useful. So Liu can agree with Batterman that phase transitions should be represented as mathematical discontinuities. Nonetheless, Liu cannot agree with Batterman that phase transitions are physical discontinuities; for, like Callender, Liu motivates his error-theory by appealing to an argument according to which, if phase transitions are represented as singularities in partition functions, there cannot be phase transitions in real systems, because partition functions for real systems cannot contain singularities.

If either Callender or Liu were to respond to Batterman's argument, they would probably challenge the claim that phase transitions are physical discontinuities.[84] For they have an independent argument to the effect that only systems in the thermodynamic limit could develop physical discontinuities.

Callender, as an error-theorist, would most likely take Batterman's argument to show that the term 'phase' needs to be redefined (or replaced with a new term, 'phase*'), so that a phase (or 'phase*') is "any homogeneous and *approximately* physically distinct part of a system which is separated from other parts of the system by a *drastic but indefinite* boundary." This modified definition does not treat different phases as qualitatively distinct. Accordingly, the modified definition is compatible with transitions between phases not being transitions between qualitatively distinct phases, and thereby compatible with phase transitions not being physical discontinuities.

Liu, as an instrumentalist, would more conservatively take Batterman's argument to show only that phases are not in fact qualitatively distinct from each other, even though it is useful, for theoretical purposes, to speak as if they are. And if phases are not qualitatively distinct from each other, there is no reason to suppose that phase

[84]There is evidence that both Callender and Liu would make this response. In a footnote, Callender assents to a statement made by Liu, that "Actual systems are finite and phase transitions in them are never real singularities" ([16], p. 550 fn. 8; [72], p. S102).

transitions are physical discontinuities, since phase transitions might just as well be sharp – albeit continuous – transitions between phases that are only *quantitatively* distinct from each other. If either Callender or Liu's response were correct, so that phase transitions are not physical discontinuities, Batterman's argument would be unsound.

5.2.2 A Direct Response

The dialectic between Callender and Liu, on the one hand, and Batterman, on the other, is indecisive. Batterman maintains that phase transitions are physical discontinuities, whereas Callender and Liu would most likely deny this. The grounds for their denial are the same as their initial grounds for advancing an error-theory about the predicate 'undergoes a phase transition'. For, if one grants that phase transitions are to be represented as singularities in partition functions, there is the purely mathematical result that only partition functions for systems in the thermodynamic limit contain singularities. And if, as Callender and Liu maintain, real systems do not exist in the thermodynamic limit because real systems have only finitely many particles, then partition functions for real systems do not contain singularities. If, as Batterman maintains, physical discontinuities are to be represented as mathematical discontinuities, then phase transitions in real systems must not be physical discontinuities; for only systems in the thermodynamic limit contain mathematical discontinuities that could represent phase transitions as physical discontinuities.[85]

[85] Batterman maintains, nonetheless, that there is something deeply correct about the thermodynamic limit. For, according to Batterman, "despite the fact that real systems are finite, our understanding of them and their behavior requires, in a very strong sense, the idealization of infinite systems and the thermodynamic limit" ([7], p. 9).

Giovanni Gallavotti suggests two constraints on any explication of 'phase transition' ([36], p. 184). First, the definition should reflect what is physically expected. For instance, one should be able to prove the existence of phase transitions for cases in which one expects phase transitions. Second, it should (hopefully) provide the tools for a closer description of typical phenomena, such as phase separation. Gallavotti's constraints illuminate the tension between Callender and Liu, on the one hand, and Batterman, on the other. The error theories proposed by Callender and Liu are designed to satisfy Gallavotti's first constraint: since the singularity-based criterion for phase transitions appears to yield the result that there are no phase transitions in real systems, and since this result does not reflect what is physically expected (contra Gallavotti's first constraint), Callender and Liu reject the singularity-based criterion. Yet, as Batterman argues, this replacement of the singularity-based criterion with a criterion about approximate singularities is not heuristically fruitful: the criteria for what counts as an approximation to a singularity are vague at best and ill-defined at worst. Moreover, if either Callender or Liu's error theory demands that phases be only *approximately* physically distinct from each other, it is not clear that the new criterion for phase transitions allows for a closer description of phenomena such as phase separation (contra Gallavotti's second constraint).

One way to advance the dialectic is to provide a direct response to Callender and Liu's error theories, a response that contests the soundness of the arguments used to establish the error theories. Their argument for an error theory regarding the predicate 'undergoes a phase transition' assumes that idealizations are false, because the argument takes the thermodynamic limit to be the limit in which a system's number of particles becomes infinite. This assumption is not mandatory. Indeed, it is possible to replace this assumption with the assumption that idealizations – and especially the thermodynamic limit – are abstractions. As the discussion from Chapter Four demonstrates, treating the thermodynamic limit as an abstracting idealization blocks the paradox of phase transitions. Since Callender and Liu's argument in favor of an error theory about 'undergoes a phase transition' is a version of that paradox, interpreting the thermodynamic limit as an abstraction also renders their argument unsound.

Since the error-theoretic arguments against the explanatory success of the statistical mechanical account of phase transitions are unsound if the thermodynamic limit is an abstraction, the burden of proof shifts to advocates of error theories. Prior to error-theoretic worries, the statistical mechanical account of phase transitions seems to be explanatory. Error theories about 'undergoes a phase transition' cast doubt upon this appearance by arguing that the account's explanandum is false. But that argument is unsound if the thermodynamic limit is an abstraction. And if the limit is an abstraction, a prima-facie case can be made to restore the plausibility of supposing that the statistical mechanical account of phase transitions is explanatory (see Chapter Four for details). The burden of proof for error theorists is to defeat that prima-facie case, without appealing to the assumption that the thermodynamic limit is a distorting idealization.

Having disposed of error theories about the predicate 'undergoes a phase transition', it remains to address error theories about the predicate 'exhibits irreversibility'. Callender's argument in favor of an error theory about this predicate assumes that any closed real system has a bounded phase space. The discussion in the previous two chapters challenges this assumption. The rigorous derivation of the Boltzmann equation shows that systems governed by the Boltzmann equation, systems that exhibit irreversible behavior, have a non-compact phase space, because the Boltzmann equation holds only of systems that are in the Boltzmann-Grad limit (see Chapter Three). Hence, if the Boltzmannian account of irreversibility is correct, some systems that exhibit irreversible behavior do not have a bounded phase space.

This alone does not challenge Callender's argument; he can still claim that the systems governed by the Boltzmann equation, the systems which exhibit irreversible behavior, are not real systems. And his claim seems to be supported by the following argument. Grant that a system exhibits irreversible behavior only if it exists in the Boltzmann-Grad limit. This is the limit in which a system's number of particles is infinite. Hence, only systems with infinitely many particles exhibit irreversible behavior. Since every real system has only finitely many particles, no real systems exhibit irreversibility.

This argument is a version of the paradox of irreversibility (see Chapter Three). The argument depends upon an interpretation of idealizations as distortions, because it treats the ineliminable idealizations that appear in the Boltzmannian account of irreversibility as distorting idealizations. This assumption is not mandatory. If the ineliminable idealizations that appear in the Boltzmannian account are abstractions, then some systems that exhibit irreversible behavior and do not have a bounded phase space are *real* systems (see Chapter Four). Treating the Boltzmann-Grad limit as an abstraction blocks the paradox of irreversibility; it also block's Callender's argument. Hence, Callender's argument in favor of an error theory about the predicate 'exhibits irreversibility' is unsound.

Since the error-theoretic arguments against the Boltzmannian account of irreversibility being explanatory are unsound if the Boltzmann-Grad limit is an abstraction, the burden of proof shifts to advocates of error theories. Prior to error-theoretic worries, the Boltzmannian account of irreversibility seems explanatory. Error theories about 'exhibits irreversibility' cast doubt upon this appearance by arguing that the account's explanandum is false. But that argument is unsound if the Boltzmann-Grad limit is an abstraction. And if the limit is an abstraction, a prima-facie case can be made to restore the plausibility of supposing that the Boltzmannian account of irreversibility is explanatory (see Chapter Four for details). The burden of proof for error theorists is to defeat that prima-facie case, without appealing to the assumption that the Boltzmann-Grad limit is a distorting idealization.

5.3 Earman's Principle

Not all criticisms of the presupposition that some explanations are ineliminably idealized are arguments in favor of some sort of error-theory about the predicates 'undergoes a phase transition' and 'exhibits irreversibility'. Nor do all such criticisms assume that idealizations are distortions. For example, John Earman claims that

> a condition of adequacy on any acceptable account of the role of idealizations [is] that it imply no effect is to be deemed a genuine physical effect if it is an artifact of idealizations in the sense that the effect disappears when the idealizations are removed ([30]).

Call this Earman's Principle. Earman's Principle is a plausible claim both under the interpretation of idealizations as distortions, and under the interpretation of idealizations as abstractions. If idealizations are distortions, then they incorrectly represent features of real systems. It seems reasonable to hold that an effect that *only* appears within a distorted model of a real system is merely an artifact of that model, rather than a feature of the real system. If idealizations are abstractions, then idealized descriptions only partially characterize real systems. Again, it seems reasonable to hold that an effect that *only* appears within a partial model of a system is merely an artifact of that model rather than a feature of the real system.

If Earman's Principle is correct, there are no ineliminably idealized explanations of genuine physical phenomena. For if there were such an explanation, then its explanandum is a phenomenon that disappears when certain idealizations are removed. Hence, according to Earman's Principle, such a phenomenon is not a genuine physical phenomenon. This is just to say that such a phenomenon does not obtain in real systems. Since there can be no explanation of a phenomenon that does not obtain in real systems, there are no explanations of phenomena that disappear upon the removal of idealizations. Consequently, there are no ineliminably idealized explanations of genuine physical phenomena.

Liu and Emch, discussing what is here called Earman's Principle in the context of explanations of quantum spontaneous symmetry breaking, argue that the principle is false. According to Liu and Emch, an important function of idealization is "to help discover (or create) – via introducing new predicates – *qualitatively* distinct properties or kinds out of ones that differ only quantitatively, more or less, prior to the idealization" ([74], p. 155). Liu and Emch claim that the different phases of a system are qualitatively distinct from each other – for example, there is a qualitative difference between the paramagnetic phase and ferromagnetic phase of a metal. This qualitative difference is best captured by appealing to the idealization of the thermodynamic limit: without that idealization, phases are represented as only quantitatively distinct. Hence, although the qualitative difference between different phases of a system disappears upon removal of the thermodynamic limit, the difference is genuine. So Earman's Principle is false.

Liu and Emch's argument against Earman's Principle is reminiscent of Batterman's argument against error-theories about the predicate 'undergoes a phase transition'. For both arguments appeal to the purported fact that the different phases of a system are qualitatively (and not merely quantitatively) distinct from each other. The advocate of Earman's Principle is thus sure to insist, with error-theorists, that the different phases of a system are not qualitatively distinct from each other. This can be insisted upon even if one agrees, with Liu and Emch, that the "difference between two phases, e.g. solid and liquid, is better captured by a singularity" ([74], p. 155), because one might hold an instrumentalist error-theory according to which, although the difference between phases is not qualitative, it is nonetheless best represented as a qualitative difference for various theoretical reasons. So, for instance, an error-theorist of an instrumentalist bent can agree with Liu and Emch that "taking the macroscopic [thermodynamic] limit is no more radical or implausible than taking space and/or time as continua; without these idealizations, the usual real analysis would not be applicable and hence many physical situations can neither be rigorously described nor inferred" ([74], p. 156). These reasons given by Liu and Emch are reasons in favor of *representing* phase transitions as singularities, rather than reasons in favor of there *being* a *qualitative* difference between different phases of a system.

Liu and Emch's argument against Earman's Principle has the same limitations as Batterman's argument against error-theories. Both arguments require the premise that differences between phases are qualitative; yet there appears to be a cogent argument that such differences cannot be qualitative for real systems, because real systems do not exist in the thermodynamic limit and only systems in the thermodynamic limit are capable of developing qualitatively different phases (because only systems in the thermodynamic limit can have singularities in their partition functions). If the Liu-Emch argument is to succeed, it must interpret the thermodynamic limit as an abstraction, in order to avoid this argument.

Although the argument against Earman's Principle given by Emch and Liu is intriguing, it is also contentious, since an advocate of Earman's Principle most likely would respond by advocating some sort of error theory about the predicate 'undergoes a phase transition'. There is a way to diffuse the force of Earman's Principle without entering into that debate. For, as a premise in an argument against the existence of ineliminably idealized explanations, Earman's Principle is question-begging.

Both advocates and opponents of the existence of ineliminably idealized explanations grant that every explanation requires that its explanandum phenomenon obtain in some real system. If P is the explanandum phenomenon of an explanation, then P occurs in some real system; and if that explanation is ineliminably idealized, explaining the occurrence of P in that real system requires an appeal to an idealization. Yet if Earman's Principle is correct, no explanation of a phenomenon that occurs in a real system requires an appeal to an idealization. For if no genuine physical effect disappears when idealizations are removed, then no genuine physical effect requires an appeal to idealization in order to be explained.

For example, according to the Boltzmannian account of irreversibility, irreversible behavior "disappears" for systems not in the Boltzmann-Grad limit. Hence, according to Earman's Principle, irreversible behavior is not a feature of real systems. Yet if the Boltzmannian account is explanatory despite being ineliminably idealized, then irreversible behavior is a feature of some real systems. For, given the assumption that the account is explanatory, and the common assumption that explanandum events obtain in some real systems, it follows that irreversibility obtains in some real systems. Moreover, claims about the reasonableness of taking features that only appear in idealized versions of systems to be non-genuine seem to be trumped by the prima-facie case that the Boltzmannian account of irreversibility is explanatory. Hence, Earman's Principle lacks dialectical force against the claim that some explanations are ineliminably idealized, because it turns out to be a straightforward denial of that claim and there is no further independent motivation for Earman's Principle.

Moreover, it is reasonable to reject Earman's Principle rather than reject the claim that some explanations are ineliminably idealized. There is a prima-facie case that the statistical mechanical account of phase transitions and the Boltzmannian account of irreversibility are explanatory. There is no extant defeater to this case. Earman's Principle seems to be motivated by the requirement that the explanans of any genuine explanation must be true; but this requirement of factual correctness can be met despite there being a violation of Earman's Principle, if idealizations are abstractions: if the description of a genuine physical effect is ineliminably idealized and the idealizations used to obtain that description are abstractions, then the description is true (albeit incomplete) even though the effect cannot be described without appealing to idealizations. Since (apparently) there is no further motivation for Earman's Principle, its threat is defused.

5.4 Emergence

Although the appeal to Earman's Principle fails to refute the presupposition that some explanations are ineliminably idealized, there remains the mystery of why there are ineliminably idealized explanations. For instance, if the free energy function that takes into account boundary and surface effects in a system cannot describe the occurrence of phase transitions in that system, why should ignoring (or otherwise idealizing) those effects allow for a description of phase transitions in the system? What is it about such phenomena that makes them immune, in principle, to non-idealized description? The short answer to this question is that phenomena like phase transitions and irreversible behavior are emergent. This section of the chapter elaborates upon this answer.

5.4.1 Constructionism

Phase transitions and irreversible behavior are phenomena that supervene upon certain facts about the systems in which they occur. For instance, the phenomenon of a system undergoing a phase transition supervenes upon facts that determine the free energy per particle for the system, such as facts about the arrangement and interactions among the constituents of the system. Likewise, the phenomenon of a system exhibiting irreversible behavior supervenes upon facts about the positions and momenta of the constituents of the system, the interactions among these constituents, and so on. These facts – the supervenience bases for the phenomena – are, in some sense, more fundamental than the phenomena that supervene upon them. Whether a system undergoes a phase transition or exhibits irreversible behavior is determined by the facts in the supervenience base for such phenomena. But the reverse is not the case; the determination relation is asymmetric.

This supervenience relation between phenomena that require an appeal to idealization for their explanation, on the one hand, and the facts that constitute the supervenience bases for these phenomena, on the other hand, generates a puzzle. The conviction that it is possible to begin from the laws and entities postulated by some fundamental theory and reconstruct the entire universe might lead one to assume the correctness of what P.W. Anderson calls Constructionism (see [3]):

> *Constructionism*: Any correct description of the facts that constitute a supervenience base for a phenomenon is thereby a correct description of the phenomenon itself.

According to Constructionism, a correct description of a supervenience base for a system suffices for a correct description of every property of the system.

Constructionism is inconsistent with the claim that the explanations (and correct descriptions) of some phenomena are ineliminably idealized. Since correct descriptions of such phenomena must be idealized, and since these idealizations pertain to facts about the supervenience bases of the phenomena, correct descriptions of these phenomena fail describe at least one fact in the supervenience base for each phenomena. The failure to describe correctly all of these facts might be due to the idealization resulting in an incorrect description of some of those facts (if idealizations are distortions); or the failure might be due to the ineliminable idealization resulting in an incomplete description of those facts (if idealizations are abstractions). For instance, a correct description of the occurrence of a phase transition in some system must be idealized. Since the required idealizations pertain to facts that determine the free energy per particle of the system (e.g., facts about what happens near the boundaries and surface(s) of the system), a correct description of the occurrence of a phase transition does not correctly describe all of the facts in the supervenience base for the occurrence of phase transitions in that system. A correct description of the occurrence of a phase transition in a system either incorrectly describes the boundary and surface effects in the system (if idealizations are distortions) or else does not describe those effects at all (if idealizations are abstractions).

According to Constructionism, if a phenomenon supervenes upon certain facts about the system in which it occurs, then any correct description of the supervenience base for the phenomenon suffices for a correct description of the phenomenon itself. However, if the only way to describe a phenomenon correctly is to appeal to some idealization of the supervenience base for that phenomenon, then any correct description of that phenomenon does not correctly describe the supervenience base for the phenomenon. Hence, if the correct description of a phenomenon is ineliminably idealized, a correct description of the supervenience base for the phenomenon does not suffice for a correct description of the phenomenon itself, contra Constructionism.

There are several ways to avoid this inconsistency. First, one might deny that phenomena like phase transitions and irreversible behavior supervene upon certain facts about the systems in which they occur. Secondly, one might deny the possibility of there being a correct description of the supervenience base for such phenomena. Thirdly, one might deny the existence of phenomena like phase transitions and irreversible behavior. Fourthly, one might deny Constructionism.

The first two of these options – denying that phenomena like phase transitions and irreversible behavior supervene upon certain facts about the systems in which they occur or denying that it is possible for there to be a correct description of the supervenience base for such phenomena – are implausible. For there is a general agreement that such a supervenience relation obtains; and, as Elliot Sober argues ([112], p. 167), the thesis that such a supervenience relation exists fits available data well enough for it to be a mistake to adopt a more complex thesis that denies supervenience.[86] Moreover, there is a general agreement that a description of a supervenience base for these phenomena is possible in principle even if impossible in practice. Given the implausibility of these two options, one must either deny the existence of phenomena like phase transitions and irreversible behavior or reject Constructionism.

To deny the existence of phenomena like phase transitions and irreversible behavior is, in effect, to adopt some sort of error theory about the predicates 'undergoes a phase transition' and 'exhibits irreversible behavior'. For if there are no such phenomena, no real system exemplifies the properties denoted by these predicates. An appeal to Constructionism thereby provides another argument in favor of an error theory about these predicates; unlike the arguments given by Callender and Liu, this argument does not assume that idealizations are distortions. Nonetheless, an appeal to Constructionism does not defeat the prima-facie case in favor of supposing that the ineliminably idealized accounts of phase transitions and irreversibility are genuinely explanatory, because the assumption that Constructionism is true is no more

[86] Michael Silberstein and John McGeever hold that emergent properties are properties of whole systems that fail to supervene upon the system constituents; they think this is "the most interesting and important kind of emergence" ([107], p. 183). Silberstein and McGeever thereby disagree with Jaegwon Kim, who "expects most emergentists to accept mereological supervenience", i.e., the supervenience of emergent properties upon their system constituents ([55], p. 7). If Silberstein and McGeever are correct and mereological supervenience fails for emergent properties, then there is no inconsistency between Constructionism and the claim that the explanations (and correct descriptions) of some phenomena are ineliminably idealized, because these phenomena fail to supervene upon their supervenience bases.

plausible than the assumption that phase transitions and irreversible behavior are emergent.

Constructionism is false if these phenomena are emergent, because a correct description of the supervenience base for emergent phenomena does not suffice for a correct description of the emergent phenomena themselves. The remainder of this section elaborates upon what it means to say that phenomena like phase transitions and irreversibility are emergent, and rebuts various objections to the possibility of there being emergent properties. If there are no plausible objections to the claim that phenomena like phase transitions and irreversibility are emergent, then the appeal to Constructionism fails to show that such phenomena do not obtain in real systems.

5.4.2 Ontological and Epistemological Emergence

There is no dearth of definitions for the notion of emergence. For the purposes of this chapter, it suffices to focus on two broad kinds of emergence, ontological and epistemological. Drawing this distinction requires first introducing the notion of an emergent predicate. Following this is a presentation of the case in favor of supposing that predicates like 'undergoes a phase transition' and 'exhibits irreversible behavior' – predicates used to describe the explananda for ineliminably idealized explanations – are emergent predicates in the ontological sense of emergence. (The discussion to follow relies heavily upon [42], pp. 55-56.)

An emergent predicate has two distinctive properties. First, it is predicable only of a system as a whole. For instance, the predicate 'undergoes a phase transition' is predicable only of a whole system; it makes no sense to say that an individual particle in a pot of water is undergoing a phase transition. (This is so, even though it *does* make sense to say that the Helmholtz free energy per particle for a system exhibits a singularity.) Likewise, the predicate ‘exhibits irreversible behavior’ is predicable only of whole systems; it makes no sense to say that an individual molecule in a gas is behaving irreversibly.

Second, the correct use of an emergent predicate allows for predictions that are impossible to derive from the relevant laws plus initial and boundary conditions alone. For instance, (proper) use of the predicate ‘undergoes a phase transition’ allows for predictions that are impossible to derive from a microscopic description of, say, a ferromagnet and the laws that govern ferromagnets, because correctly using the predicate ‘undergoes a phase transition’ requires that one consider the ferromagnet in the thermodynamic limit and the only way to predict that a ferromagnet will undergo a phase transition at a certain temperature is to consider the ferromagnet in that limit. Likewise, (proper) use of the predicate ‘exhibits irreversible behavior’ allows for predictions that are impossible to derive from a microscopic description of, say, a rarefied gas, because (properly) using the predicate requires that one consider the gas in the Boltzmann-Grad limit and the only way to predict that a gas behaves irreversibly is to consider the gas in that limit (due to the recurrence paradox).

Predicates like ‘undergoes a phase transition’ and ‘exhibits irreversible behavior’ are emergent predicates. There are two broad explanations for why emergent predicates are distinct from non-emergent predicates like ‘has mass’ and ‘is in uniform motion’. The ontological explanation is that emergent predicates pick out emergent properties. This sort of explanation results in an ontological notion of emergence, according to which there are emergent properties that are just as real as non-emergent properties such as mass.

This ontological sense of emergence accords with Jaegwon Kim's contrast between emergent properties and merely resultant ones. According to Kim,

> . . . resultant properties are . . . those that are predictable from a system's total microstructural property [i.e., the intrinsic properties of a system's particles and relations that configure those particles into a structure that is united and stable as a system], but emergent properties are those that are not so predictable ([55], pp. 7-8).

In saying that an emergent properties is not predictable from a system's total microstructural property, Kim means that "we may know all that can be known about [the supervenience base for the emergent property] – in particular, the laws that govern the entities, properties and relations constitutive of [that supervenience basis] – but this knowledge does not suffice to yield a prediction [of the emergent property]" ([55], p. 8). If emergent predicates refer to emergent properties, properties that are not predictable from total knowledge of their supervenience bases, then it is to be expected that the correct use of emergent predicates allows for predictions that are impossible to derive from the laws that govern such supervenience bases plus initial and boundary conditions alone.

In contrast to the ontological explanation for why emergent predicates differ from non-emergent ones, an epistemological explanation of this difference is that emergent predicates indicate something about our epistemic status with respect to the world. For instance, instrumentalism about the predicate 'undergoes a phase transition' distinguishes this predicate from others on the grounds that its use serves an instrumental purpose that is not readily discharged in some other way. This sort of explanation results in an epistemological notion of emergence, a notion which is non-committal about whether there are emergent properties.

There is a prima-facie case that phase transitions and irreversible behavior are emergent properties. The accounts of these phenomena appear to be explanatory despite being ineliminably idealized. Our observations seem to confirm that, at least sometimes, phase transitions occur and systems behave irreversibly. And the objections to there being such properties, based upon Callender and Liu's arguments, an appeal to Earman's Principle, or an appeal to Constructionism, are inconclusive.

5.4.3 A Defense of Ontological Emergence

Although there is a prima-facie case in favor of supposing that phase transitions and irreversible behavior are emergent properties of some real systems, there are also general philosophical objections to the effect that there are no emergent properties. The two main objections to the existence of emergent properties are due to Stephen Pepper ([91]) and Jaegwon Kim ([55]). The gist of both arguments is that emergent properties are metaphysically otiose and thereby dispensable, because they are epiphenomenal.[87] Lest the prima-facie case for the emergence of phase transitions

[87] There is a third and more recent criticism, due to Daniel Heard([42]), according to which the claim that some properties are emergent entails a highly implausible ontology. Heard's argument depends upon emergent predicates being predicates that are predicable only of systems as a whole and that yield predictions that would be very different or impossible to derive from relevant dynamical laws plus boundary conditions. Since this chapter's characterization of emergent predicates is more stringent than Heard's characterization, his arguments fail against an ontological explanation of the difference between emergent and non-emergent predicates. For instance, according to Heard's notion of an emergent predicate, the predicate 'instantiates the Central Limit Theorem' counts as an emergent predicate. This predicate does not refer to an emergent property, because it is defined as a mathematical operation on properties in the supervenience base of various samples. (If, for example, the noise voltages in a set of communication circuits are normally distributed, the set of circuits instantiates the central limit theorem.) However, the predicate 'instantiates the Central Limit Theorem' does not count as an emergent predicate according to this chapter's characterization of emergent predicates, because predictions using this predicate can be derived from relevant laws and the definition of the predicate. (For instance, one can analyze the distribution of noise voltages in a set of circuits to determine whether the distribution is normal.) Heard's argument thereby fails to apply to this chapter's claim that emergent predicates refer to emergent properties, owing to the discrepancy between this chapter's characterization of emergent predicates and Heard's characterization.

and irreversible behavior be defeated, it is necessary to show that these objections can be met. (Although I show how responses to these objections might go, I leave a proper development of the details as a further project.)

Pepper on Emergent Properties

Pepper's argument against the existence of emergent properties begins with the observation that, if an emergent property is genuine rather than merely epiphenomenal, then there must be a difference between situations in which such a property obtains and situations in which the property does not obtain. This difference must be more than a difference between the presence or absence of the emergent property itself. Hence, assuming that emergent properties are law-governed, the difference between situations in which an emergent property obtains and situations in which it does not must amount to a difference in the laws that govern such situations. This result accords well with the difference between situations that exhibit irreversible behavior and those that do not: the former are governed by something like the second law of thermodynamics, the latter are not.

Pepper's argument takes the form of a dilemma. Consider a law that governs systems in which a putatively emergent property obtains. Either this law is a primitive law, governing the behavior of the elements of the supervenience base for the emergent property; or the law is derivable from such primitive laws; or the law is a law for an epiphenomenon. There is no fourth alternative: if the law is neither primitive, derivable from primitive laws, nor a law for an epiphenomenon – if the law is "going to step down out of an epiphenomenal heaven" – then it is "bound soon to get into conflict with" the primitive laws ([91], p. 244). For there is bound to be a situation in which the emergent property obtains, such that the primitive laws governing the situation predict one behavior but the law for the emergent property predicts a different behavior.

Given these three alternatives for the kind of law the emergent law might be, it follows that either the putative emergent property governed by the law in question is not emergent or the putative emergent property governed by the law is emergent but epiphenomenal. For if the law that governs systems in which the putative emergent property obtains is primitive or derivable from primitive laws, then the property it governs is not emergent, in virtue of being predictable from the supervenience base for the property. And if the law that governs systems in which the putative emergent property obtains is a law for an epiphenomenon, then of course the emergent property is an epiphenomenal property.

Pepper takes his argument to refute the existence of emergent properties, because he assumes that "a theory of wholesale epiphenomenalism is metaphysically unsatisfactory" ([91], p. 241). A quick response to Pepper's argument would be to bite the bullet or deny his metaphysical intuitions, accepting that emergent properties are epiphenomenal. But there is a better response available. Pepper overlooks a way to ensure consistency between primitive laws and laws for genuine, non-epiphenomenal emergent properties. For, in addition to the alternatives he considers, there is the possibility that the primitive laws have a restricted range of validity, that the laws "break down" when applied to systems in which genuinely emergent properties obtain. (Paul Meehl and Wilfrid Sellars raise this possibility in [80].)

If the primitive laws break down when applied to systems in which genuinely emergent properties obtain, then there is no situation in which emergent laws get into conflict with the primitive laws. And a case can be made that the primitive laws do break down in this way. In a different context, Alexander Rueger suggests that the relation between instances of emergent properties and their supervenience bases is a part-whole relation: an instance of an emergent property in some system is "part" of instances of the properties of the supervenience base for that instance in that system (see [100], p. 13). Following this line of thought, the difference between primitive laws and emergent laws is that primitive laws apply to systems as a whole, whereas emergent laws apply to mere "parts" of those systems. And the reason why primitive laws break down when applied to systems in which genuinely emergent properties obtain, is that the primitive laws account for "too much other stuff", stuff that is irrelevant to the obtaining of the emergent properties. This other stuff "obscures" the emergent properties. The emergent laws ignore this excess baggage, which is why they are able to account for the obtaining of emergent properties.

Admittedly, this response to Pepper's argument raises many issues: what is a "part" of a system? how can properties be "obscured" by laws that take into account too much detail? to what extent is this obscuring a pragmatic function of our interests, and to what extent is the obscuring metaphysical? Satisfactorily resolving these issues is a project in itself. Here it suffices to indicate that Pepper's argument can be avoided, without providing all of the minute details for the way in which this can be accomplished.

Kim on Emergent Properties

Jaegwon Kim provides a second argument to the effect that emergent properties must be epiphenomenal. Kim's argument shows the inconsistency of the claim that emergent properties are *not* epiphenomenal with five other plausible claims. These other claims are:

1. Supervenience: Emergent properties supervene upon primitively physical (non-emergent) properties.

2. Distinctness: Emergent properties are wholly distinct from primitively physical properties.

3. Causal Closure: Every primitively physical event (involving only primitively physical properties) that has a sufficient cause at a time t has a sufficient primitively physical cause at t.

4. Downwards Causation: If an event e_2 supervenes on an event e_3 and if e_1 is a sufficient cause of e_2 at time t, then e_1 causes e_2 in virtue of being a sufficient cause of e_3 at t.

5. Causal Exclusion: If an event e_2 has a sufficient cause e_1 at t, then there is no other event wholly distinct from e_1 that is also a sufficient cause of e_2 at t.

Causal Exclusion entails that if the collision of a brick with a window is a sufficient cause of the window's breaking, there is no other event that is also a sufficient cause of the window's breaking.

Kim's argument proceeds as follows. For reductio, suppose that some emergent event E_1 is a sufficient cause of an emergent event E_2 at time t. Since the emergent supervenes on the primitively physical, E_2 supervenes on some primitively physical event P_2. The event E_1 causes E_2 at t in virtue of being a sufficient cause of P_2 at t, via Downward Causation. Since the primitively physical is causally closed, there is also a sufficient primitively physical cause of P_2 at t (say, P_1). This primitively physical cause is wholly distinct from E_1, since the emergent and the primitively physical are distinct. Hence, via Causal Exclusion, E_1 is not a sufficient cause of E_2 at t. Contradiction.

If one assumes the metaphysical unsatisfactoriness of any theory according to which emergent properties are epiphenomenal, Kim's argument refutes the existence of emergent properties. A quick response to this argument would be to bite the bullet and deny the metaphysical intuition, accepting that emergent properties are epiphenomenal. One might also defend the view according to which emergent events are systematically overdetermined, in which case Causal Exclusion is false and Kim's argument is unsound. Or, following the response to Pepper's argument, one might hold that instances of emergent properties are "parts" of instances of the properties of the supervenience base for those instances. In this case, the Distinctness premise is false and Kim's argument is unsound (see [100], p. 13).

5.5 Conclusion

There are three kinds of objection to the presumption that some explanations are ineliminably idealized: error-theoretic; those based upon Earman's Principle; and those based upon an appeal to Constructionism. All of these objections can be met. The error-theoretic objections fail if idealizations are abstractions. Objections based upon Earman's Principle can be shown to be question-begging. And objections based upon Constructionism can be avoided by taking the explananda of ineliminably idealized explanations to be ontologically emergent phenomena.

CHAPTER 6

IDEALIZED EXPLANATIONS AS ONTOLOGICAL GUIDES

Inference to the best explanation is a scheme of inference for selecting which hypothesis, from a set of incompatible hypotheses, is most probably true. When cogent, inference to the best explanation is also a route that connects explanation and ontology. Most scientific hypotheses, however, happen to be idealized; and if idealizations are false, inference to the best explanation is not a cogent form of inference. The question thus arises: if two idealized hypotheses, both explanatory, are incompatible with each other, which – if either – is a guide to what the world is like?

This chapter has two aims. The first is critically to discuss two philosophical accounts of the connection between idealized hypotheses and ontology. These accounts are due to Lawrence Sklar and Paul Teller, and they share the assumption that idealizations are false. One thesis of this chapter is that neither of these accounts adequately characterizes the connection between idealized hypotheses and ontology. The second aim of this chapter is to present an alternative account of this connection. The account to be presented rejects the assumption that idealizations are false, in favor of the assumption that idealizations are abstractions (in the sense discussed in Chapter Four). The second thesis of this chapter is that the resulting account more adequately characterizes the connection between idealized hypotheses and ontology, than do extant accounts that take idealizations to be false. This provides further support for an interpretation of idealizations as abstractions, by showing that the distinction between distortions and abstractions can solve a problem other than the one of how ineliminably idealized accounts can be explanatory. This chapter shows that treating idealizations as abstractions rather than as distortions is also useful in cases that involve eliminable idealizations, since the idealized hypotheses discussed herein are eliminably idealized.

6.1 Explanation, Ontology, and Idealization

Science is rife with hypotheses that, although potentially explanatory, are incompatible with each other. There is the incompatibility between Darwin's theory of evolution and a once-accepted natural theology, two hypotheses that seek to explain the origin and diversity of species. There is also the incompatibility between Lavoisier's theory of oxygen and a once-accepted theory postulating the existence of phlogiston, two theories that seek to explain combustion and the calcination of metals. Thirdly, there is the incompatibility between the wave theory and particle theory, two hypotheses that seek to explain the behavior of light. Again, there is the incompatibility between Newtonian mechanics and the general theory of relativity, two hypotheses that seek to explain the behavior of celestial objects.[88]

Within physics, often one finds idealized hypotheses about the structure or constitution of a physical system that, while potentially explanatory, are incompatible with each other. For example, the gross dynamical behavior of a metal object spinning in

[88]These examples are taken from [119].

a force field can be explained by characterizing the object as a perfectly rigid (non-distortable) body, while the way in which changes in the force field distort this same object can be explained by characterizing the object as a body of discrete elements joined with binding forces (see [110], p. 430). For another example, consider the liquid drop model of the nucleus, which treats the nucleus as an incompressible liquid and explains nuclear deformation and fission, while the shell model of the nucleus treats the nucleus as a collection of discrete nucleons and explains nuclear binding energies and phenomena for which spin is important.

It is clear that most of our hypotheses about the world are idealized in some way. Hence, although idealized hypotheses are not entirely true of the systems they characterize, they are our best guides to what those systems are like. (In saying that an hypothesis is a guide to what a system is like, I mean that the hypothesis provides a characterization of the system that is to be endorsed as true until a better one is available.) However, if a physical system is usefully characterized by different and incompatible idealized hypotheses, at most one can be treated as a guide to the ontology of the system. For example, if a metal object is usefully characterized as both a perfectly rigid body and a collection of discrete elements joined by binding forces, at most one of these characterizations can be taken as a guide to what the object is actually like, lest the object be taken to be both rigid and non-rigid. Eschewing such absurdity, either apparently incompatible idealized hypotheses about the same system are not actually incompatible, or only some idealized hypotheses are ontological guides.

A natural way to decide which idealized hypothesis, from a set of competing explanatory hypotheses, is an ontological guide, is to invoke inference to the best explanation.[89] Inference to the best explanation is a scheme of inference for selecting which hypothesis, from a set of incompatible hypotheses, is most probably true. According to inference to the best explanation, that hypothesis with the most explanatory power is to be taken as the hypothesis that is most probably true.[90] Superiority in explanatory power is, arguably, the reason why Darwin's theory is to be accepted rather than natural theology, and why the oxygen theory is to be accepted rather than the phlogiston theory.[91]

On the supposition that inference to the best explanation is a cogent form of inference,[92] it sometimes connects explanation and ontology. Inferences to the best explanation conform to a general pattern:

1. Hypothesis H is a potential explanation of a set of data D about some set of phenomena: if H were true, it would explain D.

2. H is the best potential explanation of D from among the currently available explanations.

[89] Another way is to appeal to Bayesian confirmation theory, taking the hypothesis with the highest probability to be an ontological guide. Michael Shaffer has argued that Bayesianism cannot accommodate hypotheses that are idealized; see [105]. So it is far from clear that a Bayesian strategy is applicable here.

[90] The assumption throughout this chapter will be that if an hypothesis is potentially explanatory, the potential explanation it provides is "good enough" for it to be at least a potential guide to what the world is like.

[91] Of course, there is also the fact that there is no evidence for the existence of phlogiston. But the superior explanatory power of oxygen theory over phlogiston theory is at least part of the reason for preferring oxygen theory to phlogiston theory: even without evidence for the existence of oxygen, its superior explanatory power privileges it as the hypothesis to be accepted (rather than phlogiston theory).

[92] This is a contestable – and contested – supposition. For criticism, see [122]. For a reply, see [94]. Addressing this issue is beyond the scope of this chapter.

3. Therefore, H is probably true.

When the hypothesis involved in an inference to the best explanation concerns the structure or constitution of a physical system, inference to the best explanation is a way of inferring an ontology of the world: if the hypothesis is the best explanation, then the world is probably structured or constituted the way the hypothesis says it is. For instance, if the wave theory is a better explanation of optical phenomena than the particle theory (and otherwise a "good enough" explanation), then light is probably wave-like rather than particle-like. Similarly, if the shell model is a better explanation of nuclear phenomena than the liquid drop model (and otherwise a "good enough" explanation), then the nucleus is probably a collection of discrete nucleons rather than an incompressible liquid. Since only one of multiple incompatible idealized hypotheses about a system can be the best explanation of a set of phenomena, privileging the best explanation as ontological guide guarantees a consistent theory of what the world is like. (Of course, two hypotheses might tie for having the most explanatory power with respect to some set of phenomena; in this case, I prefer to say that neither is *the* best explanation and that neither serves as an ontological guide. In saying that an hypothesis is the best explanation of some set of phenomena, I mean that it has more explanatory power than any of its current competitors.)

The problem with using inference to the best explanation to select an ontological guide from a set of competing idealized hypotheses is that the cogency of this inference depends upon the way in which idealizations are interpreted. As a distortion, an idealization of some property is a false characterization of that property; and an idealized hypothesis about a system is an hypothesis that is false of the system. If idealizations are distortions, then since any idealized hypothesis that furnishes an idealized explanation is known to be false, any inference to the conclusion that the hypothesis is probably true is not a cogent inference. Inference to the best idealized explanation is not cogent if idealizations are distortions.

A putative aim of science is to discover what the world is like. Since most scientific hypotheses are idealized, idealized hypotheses are often the only route available for inferring an ontology that suffices until something better comes along. Given the assumption that idealizations are distortions, there must be a way to decide which hypothesis, from a group of incompatible idealized hypotheses, is an ontological guide in a way that does not appeal to explanatory considerations but nonetheless avoids ontological inconsistency.

There are two extant accounts of the connection between idealized explanations and ontology, both of which interpret idealizations as distortions. According to Lawrence Sklar, explanatory idealized hypotheses that are “on the road to truth” are guides to ontology, while others are “convenient fictions.” According to Paul Teller, *every* explanatory idealized hypothesis is a guide to ontology, but this does not result in an inconsistent theory of what the world is like because incompatible idealized hypotheses about the same system characterize different aspects of the system – each hypothesis provides a different perspective on the same system. The aim of this chapter is to show that both of these accounts are unsuccessful, and to show that an account that rejects the interpretation of idealizations as distortions is able to provide a satisfactory account of the connection between idealized explanations and ontology.

6.2 Distorted Hypotheses as Ontological Guides

If idealizations are distortions, then idealized hypotheses are known to be false. In virtue of their being idealized, it can be known that idealized hypotheses provide an incorrect description of what the world is like. Nonetheless, since most of our hypotheses are idealized, these incorrect descriptions are the only route available for inferring a provisional ontology, an ontology that suffices until something better comes along. But what is it about some idealized hypotheses that privileges them as ontological guides?

This question is especially pertinent when there are incompatible idealized hypotheses that characterize the same system. Consider, for instance, the incompatible idealized hypotheses about nuclear structure. If the shell model is taken to be an ontological guide, then the nucleus is (provisionally) a collection of discrete nucleons. If the liquid drop model is taken to be an ontological guide, then the nucleus is (provisionally) an incompressible liquid continuum. But nothing can be both discrete and a continuum at the same time, in the same respect. Since both models are idealized and idealizations are being interpreted as distortions, inference to the best explanation cannot decide which model (if either) to privilege as the ontological guide.

6.2.1 Alethic Trajectories and Interest-Relativity

According to Lawrence Sklar, what privileges an idealized hypothesis as a guide to ontology is its being "on the road to truth." Sklar distinguishes between hypotheses that are "on the road to truth" and those that are merely "convenient fictions" and hence ontologically erroneous. According to Sklar,

> Multiple incompatible schemes applied to one and the same system, each one of which has a legitimate and explanatory use, cannot all be intended

> to be genuinely "true" of the system. We must allow for "convenient fictions" . . . ([110], p. 438).

Idealized hypotheses that are "convenient fictions" might be explanatory, but they are not guides to inferring what the world is like. Being an ontological guide is a privilege for those hypotheses which, although false and known to be so, are nonetheless "on the right track". (Sklar does not develop his metaphors; but what he says seems to entail that at most one hypothesis, from amongst a group of competing hypotheses, can be "on the right track", because at most one can be an ontological guide.)

Whether an hypothesis is "on the right track" or "on the road to truth" is to be decided by attention to the details of scientific practice ([110], pp. 431, 439). Evidence that "there is at least some domain of physical situations for which [an hypothesis] will remain a reliable predictor of observational outcomes into the perpetual future" is evidence that the hypothesis is on the right track (see [109], p. 89). For example,

> there is something very different between characterizing an atomic nucleus as a complex system of neutrons and protons, with these compounds composed of quarks bound by gluons, and with the neutrons and protons bound by a van der Waals residual effect of the quark-quark binding [shell model], and a characterization of a fissionable nucleus as a "liquid drop" held together by a "surface tension" [liquid drop model] ([110], p. 439).

Sklar takes the shell model to be "at least a part of a structure 'on the road' to our desired ultimate theory", and the liquid drop model to be a convenient fiction, "a weak model adequate only in the most restricted ways to characterizing what is really going on." Presumably this is because the shell model hypothesizes a nuclear structure that tightly coheres with the ontology postulated by the Standard Model, while the ontology of the liquid drop model coheres less tightly. Although the Standard Model is itself idealized, its predictive success ensures that, in the future, it will remain a reliable theory in some appropriately restricted (but as yet unknown) domain. Hence, although both the shell model and the liquid drop model are known to be false, the shell model is "on the right track." The shell model is an ontological guide while the liquid drop model, as a merely convenient fiction, is not.

Paul Teller objects to Sklar's account, on the grounds that Sklar's distinction, between hypotheses that are on the right track and those that are convenient fictions, is irrelevant to privileging the former as ontological guides rather than the latter. Teller points out that the entities postulated by the Standard Model, quarks and gluons, are excitations of the positive and negative frequency solutions of a wave equation. Since the most accurate models of the structure of space-time hypothesize that space-time is irregularly curved, and since there are no positive and negative frequency solutions of the field equations in these models, quarks and gluons are "idealizations every bit as much as the idealization of a liquid as a continuous medium" and, one might add, every bit as much as the idealization of a nucleus as a liquid drop ([117], p. 433). So the current state of science fails to provide evidence to support the claim that the Standard Model, rather than the liquid drop model, is on the right track.

Teller further argues that any difference between two hypotheses, both of which are known to be false, must be a matter of degree. Every false hypothesis, insofar as it is explanatory, is "on the road to truth" to some degree. For, in virtue of its being explanatory, the hypothesis is getting something right, even if the hypothesis is false overall. There is something right about the liquid drop model, since it explains nuclear deformation and fission; but there is also something right about the shell model, since it explains nuclear binding energies. Likewise, there is something right about taking a spinning metal body to be perfectly rigid; but there is also something right about taking the same body to be a set of discrete elements joined by binding forces.

Moreover, Teller continues, the degree to which a false hypothesis is "on the right track" depends upon contextual interests.

> ... "closer to the truth," "more accurate," and the like make perfectly good sense, but only in a relational way, relative to things like aspects and features that in turn may be variously evaluated in relation to our interests ([117], p. 438).

For instance, perhaps the liquid drop model is "farther down the road to truth" than the shell model for those concerned with the creation of energy via nuclear fission, while other concerns reverse the situation.[93] Since nothing privileges some interests as more important than others, there is no objective measure of the degree to which two competing idealized hypotheses are on the right track, of how far down the road to truth two false hypotheses happen to be. Any such measure is interest-relative.

Yet, according to Teller, no interest-relative difference between two false hypotheses is relevant to privileging one hypothesis as an ontological guide rather than the other. For what the world is like is not interest-relative. Since the distinction between false hypotheses that are on the road to truth and those that are merely convenient fictions is interest-relative, Sklar's distinction is irrelevant to privileging the former as ontological guides rather than the latter: "the metaphor of 'farther down the

[93]Teller makes a similar point in his discussion of quantum mechanical and hydrodynamical characterizations of fluids: "... when it is the fluid properties of water that are of interest, a hydrodynamic characterization of water may be fairly evaluated as much more 'truth-like' than a quantum mechanical description, let alone any humanly accessible characterization in terms of quantum field theory" (2004, 440). Obviously, with respect to different interests, quantum mechanics provides a better characterization of certain systems than hydrodynamics.

road to truth' won't help with driving a wedge between acceptable and unacceptable ontologies" ([117], p. 434).

Perhaps it is possible to distinguish false hypotheses that are on the right track from those that are not, in a way that is not interest-relative. Certainly Teller's criticism does not rule out this possibility. Still, the burden of proof rests squarely with those who suppose that such a distinction is possible. Meeting this burden is made more difficult by the exclusion of explanatory qualities that might differentiate some hypotheses from others. For, as has been noted, inference to the best explanation is not cogent under the assumption that idealizations are distortions. Moreover, the appeal to the truthlikeness or approximate truth of a false hypothesis will not do, because, as Teller convincingly argues elsewhere ([116]), the degree to which a false hypothesis is truthlike is interest-relative.

On Sklar's behalf, one might attempt to understand what it is for an hypothesis to be "on the right track" in terms of a modified form of inference to the best explanation, arguing that the false hypothesis with the most explanatory power is the guide to what the world is like, and that such an hypothesis can serve as an ontological guide without our inferring that it is true. This would allow explanatory qualities to differentiate some false hypotheses from others without requiring an illicit appeal to inference to the best explanation.

However, such an approach will face the following sort of problem. Taking the false hypothesis to be an ontological guide, without inferring that the hypothesis is probably true, requires an attitude other than belief towards the hypothesis, since one cannot believe an hypothesis that one knows to be false. Acceptance, which is generally taken to be weaker than belief, seems to be an inappropriate attitude to have towards an hypothesis one knows to be false, because accepting an hypothesis as an ontological guide at least seems to require not knowing (or believing) that it is false.[94] And there does not seem to be any attitude one might have towards an hypothesis one knows to be false that allows one to consistently maintain that the world is the way the hypothesis characterizes it.

6.2.2 Ontological Pluralism

Sklar's account of what privileges some false hypotheses as ontological guides rather than others faces an apparently insurmountable problem of finding a criterion that is not interest-relative and that does not appeal to explanatory qualities of such hypotheses. This poses a dilemma: on the one hand, sometimes at least one hypothesis is a guide to what the system is like, since there is a presumed (albeit provisional) connection between explanatory idealized hypotheses and ontology; on the other hand, incompatible false hypotheses about the same system cannot all be guides to what the system is like. Since not every hypothesis can be an ontological guide, then there must be a non-interest-relative criterion that privileges one over the others as a guide to ontology; but there does not appear to be such a criterion. As a result, it seems that no idealized hypotheses can be ontological guides.

Teller escapes this conclusion by rejecting the assumption that two false hypotheses about the same system cannot both be guides to what that system is like. According to Teller,

> When we give up this presumption we see that we can embrace 'conflicting' ontologies, being careful not to invoke them 'at the same time,' or, more carefully, not to take them to represent the same aspects of the not

[94]Of course, if one is an anti-realist about the hypothesis in question, one might accept the hypothesis despite knowing that it is false. But, as an anti-realist, one does not maintain that the hypothesis characterizes the way the world is.

completely accurately represented objects of our representations" (2004, 441).

That is, when there are at least two false but explanatory hypotheses that characterize the same system, both are to be taken as ontological guides; but only if the hypotheses are guides to the nature of different aspects of the system. As Teller puts it, "we may take a pluralist ontology to consist of a collection of idealized descriptions that, when deployed with care in cognizance of their limitations, can be consistently applied as complementary rather than conflicting" ([117], p. 441).[95]

For example, since the shell model and liquid drop model of the nucleus are both explanatory of nuclear phenomena despite being false, both are guides to aspects of what the nucleus is like. One aspect of the nucleus is its discrete structure of nucleons; this is the aspect of the nucleus that is relevant to nuclear binding energies. A different aspect of the nucleus is its incompressibility as a liquid; this aspect is relevant to nuclear fission. There is one aspect of the nucleus according to which it is very much like a collection of discrete nucleons, there is a different aspect according to which it is very much like a continuum, and since these are different aspects of the same nucleus there is no inconsistency in the theory of what the nucleus is like. The shell and liquid drop models are complementary rather than incompatible, because they characterize different aspects of the same system.

Again, in studying a fluid it is sometimes useful to model the fluid as a continuum, in which case thermodynamics can be brought to bear upon the study. In other cases it is useful to model the fluid as a collection of discrete particles, in which case statistical mechanics can be invoked. The fluid itself, according to Teller, has one

[95]For a similar account of apparent incompatibility, in the context of biological explanations and the ontology of biological systems, see [83].

aspect that is continuum-like and a different aspect that is nicely approximated as a collection of discrete particles. The different explanatory hypotheses yield a pluralist ontology of what a fluid is like. They are complementary rather than incompatible, because they characterize different aspects of the same system.

According to Teller, if an idealized hypothesis is explanatory, it is also a guide to ontology. On this view, every idealized model that is explanatory has an equal right to be a guide to what the world is like, because there is no way to privilege some false hypotheses over others in a way that is not interest-relative. Potential conflicts among apparently incompatible hypotheses are to be avoided by taking each hypothesis to characterize a different aspect of the system of interest.

Teller claims that ontological pluralism is what we should expect from using, as ontological guides, hypotheses that are known to be deficient.

> Our representations are intended as a guide to the world. But our guides are imperfect – they speak to us fallibly. So their virtues as guides to the world cannot be simply evaluated in terms of the dichotomy, true or false. Rather a sensible objective is degree of fit in respects that are of current interest ([117], p. 439).

That is, although idealized hypotheses are explanatory, they are also limited in virtue of their being idealized; they get something right, but not everything. This is another way of saying that idealized hypotheses characterize aspects of systems, but not systems in their entirety.

One consequence of Teller's ontological pluralism is that idealized explanations are not guides to what the world is like *per se*. Rather, each is a guide to what the world is like from some perspective, each perspective providing access to a different aspect of the world. There is no "getting behind" the various perspectives, so to speak, because idealized hypotheses, in virtue of being false, at best characterize aspects of systems. Hence, idealized explanations provide no access to a non-aspectual characterization of what a system is like. In this respect, Teller's ontological pluralism is reminiscent of Bohr's notion of complementarity as a tool in an early quantum mechanical interpretation of the nature of light: light has particle-like aspects and wave-like aspects, and nothing more can be said.

The thesis of ontological plurality revises the original assumption of there being a connection between idealized explanations and what the world is like without reference to anything else. To accept ontological pluralism is to accept that every characterization of what the world is like is relative to a perspective. This is no criticism of the thesis, of course. But the consequence is worth pointing out, since an account of the connection between idealized hypotheses and ontology that does not have this consequence, but is nonetheless at least as adequate as Teller's ontological pluralism, is thereby preferable, on the basis of its being less apparently anthropocentric than Teller's account. (In relativizing ontology to perspectives, Teller seems to introduce a human-centered relativism into our ontology of the world.)[96]

There is, moreover, an additional problem with Teller's account: it presumes that whenever incompatible idealized hypotheses explain some feature of the same physical system, each characterizes a different aspect of that system. There are common violations of this presumption. Consider, for example, two incompatible idealized explanations of the rough, qualitative proportionality between a pendulum's length and its period. The simplest explanation of this proportionality is provided by appeal to the simple pendulum. Taking idealizations to be false, the simple pendulum has a point-mass bob, massless and rigid rod, frictionless pivot, etc. A different explanation

[96]I thank Bob Batterman for suggesting this point.

of the proportionality between a pendulum's period and its length is provided by appeal to an extended-bob pendulum. Taking idealizations to be false, an extended-bob pendulum has an extended bob, rigid but not necessarily massless rod, frictionless pivot, etc. If idealizations are distortions, no pendulum can be both a simple pendulum and an extended-bob pendulum; so these characterizations are incompatible if taken to apply to a pendulum in its entirety. Yet the simple pendulum is used to characterize the same aspect of real pendula as is the extended-bob pendulum, namely, the qualitative proportionality between period and length. Hence, these idealized hypotheses do not characterize different aspects of the same pendulum system. Teller's ontological pluralism provides no guidance on the connection between these idealized hypotheses about pendulum behavior and what real pendula are like.

Again, the gross dynamical behavior of a metal object spinning in a force field can be explained by characterizing the object as a perfectly rigid body. This same behavior can also explained, albeit with more complications, by characterizing the object as a body of discrete elements joined by binding forces. These two hypotheses characterize the same aspect of the spinning metal object, namely, its gross dynamical behavior. Hence they do not characterize different aspects of the object. Nonetheless, they are incompatible with each other: no object can be both perfectly rigid and also non-rigid. Teller's ontological pluralism provides no guidance on the connection between these idealized explanations and what the metal object is like.

6.3 Abstract Hypotheses as Ontological Guides

The preceding discussion suggests two constraints on a satisfactory account of the connection between idealized hypotheses and ontology. (1) If, following Sklar, the account treats some explanatory idealized hypotheses as ontological guides but treats others as merely "convenient fictions", the criteria it uses to distinguish between the guides and the fictions should not depend upon contextual interests. (Sklar's account fails to provide such criteria.) (2) Whether the account differentiates ontological guides from merely convenient fictions or – like Teller's account – does not, it should permit privileging some hypotheses over others as ontological guides when both pertain to the same aspect of the same system. (Teller's account fails to provide such permission.)

Both of these constraints can be met by treating idealized hypotheses as abstract descriptions and treating idealizations as abstractions rather than as distortions. (The notion of an abstraction is presented in Chapter Four.) If an idealization is an abstraction, it replaces one description of a system with a description that fails to attribute to the system at least one feature that the system has, without thereby attributing to the system a feature it does not have. The resultant description, an abstract description, merely omits mention of some feature that the system has; it is an incomplete description of the system. If idealizations are abstractions rather than distortions, then idealized hypotheses are incomplete but they need not be false.

The incompleteness of idealized hypotheses is compatible with the cogency of inference to the best explanation, since the incompleteness of an hypothesis is compatible with its (probable) truth. This section shows how inference to the best explanation can provide a connection between idealized hypotheses and ontology, under the assumptions that idealizations are abstractions and that inference to the best explanation is cogent.

6.3.1 Inference to the Best Idealized Explanation

If one accepts the cogency of inference to the best explanation, then not every idealized hypothesis is an ontological guide in virtue of its being explanatory: only those idealized hypotheses that provide the best explanations are guides to ontology.[97] Schematically, inference to the best idealized explanation takes the following form:

1. Idealized hypothesis H is an idealized potential explanation of data D.
2. H is the best idealized potential explanation of D.
3. Therefore, H is probably true.

The best idealized explanation is the one, from amongst a group of competing explanations, that has the most explanatory power. (Of course, the conclusion of an inference to the best explanation is defeasible.)

There is, at present, no consensus on how to judge the comparative explanatory power of competing idealized hypotheses. Nonetheless, it might be helpful to list several criteria that seem to be plausible. These criteria usually are not taken to pertain to *idealized* explanations; but they adapt readily to such a purpose. They are:

[97] Sklar writes that "Our inferences to 'best explanations,' whatever such inferences are like, is [sic], surely, our optimal inferential route for inferring our believed ontology of the world" ([110], p. 425). It is surprising that, rather than pursuing this claim, Sklar turns his attention to a distinction among alethic trajectories of models. Perhaps the reason for this is that Sklar is thinking of the data set for each inference as restricted to one phenomenon. Given this restriction, it is certainly possible to end up with an inconsistent ontology. The way to avoid this, however, is not to abandon reliance on inference to the best explanation, but rather to expand the data set to cover all phenomena for which there is some sort of explanation. (Alternatively, one might take the discussion to follow to be the proper fleshing out of Sklar's metaphoric distinction between hypotheses "on the road to truth" and merely convenient fictions.)

1. *Consilience*: If two idealized hypotheses explain the same data, but background knowledge favors one over the other and there is no specific reason to challenge the background knowledge, then the former hypothesis has more explanatory power than the latter.

2. *Completeness*: If only one idealized hypothesis, from amongst a group of idealized hypotheses, explains all the data that the hypotheses taken together explain, then it has more explanatory power than the other hypotheses.

3. *Importance*: If one idealized hypothesis explains data that are more salient than the data explained by a different idealized hypothesis, then the former has more explanatory power than the latter.

4. *Parsimony*: If two idealized hypotheses explain the same data, but one invokes auxiliary assumptions that form a proper subset of the auxiliary assumptions invoked by the other, then the former idealized hypothesis has more explanatory power than the latter.

5. *Precision*: If two idealized hypotheses explain the same data, but the idealized explanations given by one are more precise than those given by the other (in the sense that the former appeal to causal-nomological mechanisms whereas the latter do not), then the former hypothesis has more explanatory power than the latter.

These criteria are adapted from those provided by Stathis Psillos ([96]). Oftentimes, it is not clear how to rank the relative explanatory power of competing idealized hypotheses; and sometimes hypotheses are ranked differently by different criteria.

But addressing these problems is a project for some other time. The important point here is that these criteria provide a way to privilege some idealized hypotheses as ontological guides rather than others.

Unlike Sklar's distinction between idealized hypotheses "on the road to truth" and those that are merely "convenient fictions", rankings of the relative explanatory power of idealized hypotheses are not interest relative. The criteria for ranking explanatory power are intended to either preserve or enhance the explanatory coherence of the corpus of scientific beliefs. The degree to which a set of beliefs is explanatorily coherent is not interest-relative, at least according to some accounts of explanatory coherence.[98] And since explanatory coherence is not interest-relative, criteria that preserve or enhance explanatory coherence are not interest relative.

This line of argumentation entails that at least some of the preceding criteria for ranking explanatory power can be characterized in a manner that is not interest relative. No argument has been given to substantiate this corollary. Doing so is a project for some other time, especially since I am not endorsing the correctness of the preceding criteria. Nonetheless, it seems to be fairly obvious that rankings of explanatory power according to the criteria of Consilience and Completeness are not interest-relative. And these are the only criteria that play a role in the remainder of this chapter.

6.3.2 Partiality without Plurality

Given these criteria for ranking the relative explanatory power of idealized hypotheses, it is possible to connect idealized explanations and ontology in a way that avoids the interest-relativity that besets Sklar's account and in a way that avoids

[98] One such account is given by Paul Thagard; see [120].

Teller's ontological pluralism. If two or more idealized hypotheses are explanatory and yet characterize the same system, then the hypothesis that provides the best idealized explanation is the one to be privileged as an ontological guide, in accordance with inference to the best explanation.

Which hypothesis provides the best idealized explanation is a matter of which has the most explanatory power relative to its competitors. For instance, the shell model ignores features of the nucleus that are relevant to nuclear fission but not features relevant to nuclear binding energies; the liquid drop model ignores features relevant to nuclear binding energies but not relevant to nuclear fission. Although each model ignores some features of the nucleus that are relevant to some nuclear phenomena, one might argue that the shell model is a better explanation of nuclear phenomena than the liquid drop model because the shell model is more closely integrated with the Standard Model (*Consilience*). Hence, the shell model rather than the liquid drop model would be the guide to what the nucleus is like. Again, one might argue that characterizing a metal object as a collection of discrete elements provides a better explanation of the properties of that object than does characterizing it as a perfectly rigid body, because the former characterization can explain everything covered by the latter characterization, and more besides (*Completeness*). Inference to the best explanation is cogent in these cases, because the idealized hypotheses are treated as incomplete rather than false.

This account of the connection between idealized explanations and ontology respects Teller's observation that idealized hypotheses serve as guides to what the world is like only in an imperfect way. In virtue of being idealized, every idealized hypothesis is incomplete in some way or another. Yet some idealized hypotheses are less imperfect than others, in virtue of their being less idealized and thereby representing more features of the systems they characterize. (Of course, the results from preceding chapters of this dissertation show that hypotheses that represent more features of the systems they characterize are not always less imperfect than more idealized hypotheses. Rankings of the degree of imperfection of a set of hypotheses do not always correspond to rankings of the degree to which those hypotheses are idealized, because some idealizations are ineliminable to any correct description of certain phenomena. However, since the idealized hypotheses discussed in this chapter are *eliminably* idealized, the less idealized hypotheses are less imperfect.)

This account also avoids the problems raised for Teller's account. As noted, both the simple pendulum and the extended-bob pendulum can be used to provide an idealized explanation of the qualitative proportionality between a pendulum's length and its period. Taking idealizations to be abstractions, the simple pendulum does not represent the amounts for the extension of the pendulum bob, the mass or flexibility of the pendulum rod, the friction at the pivot, etc. Unlike the simple pendulum, the extended-bob pendulum represents the amount of extension of the pendulum bob; but like the simple pendulum, it does not represent amounts for the mass or flexibility of the pendulum rod, the friction at the pivot, etc.

There is no inconsistency in saying that the simple pendulum and the extended-bob pendulum characterize the same system, because if idealizations are abstractions they both can provide a partial characterization of the same real pendulum. An analogous case illustrates this point. Suppose that Melissa and Barry are describing a car's features. Melissa says that the car has four doors, a nice shine, and a moon roof; but that's all she says. Barry says the car has four doors, a nice shine, a moon roof, a V6 engine, low mileage, anti-lock breaks, etc. Barry's description is more detailed than Melissa's more partial characterization, but not thereby inconsistent with it.

Likewise, the simple and extended-bob pendula provide different characterizations of the same aspect of a real pendulum, namely, the qualitative proportionality between its period and length. The characterization by the simple pendulum ignores some features of the real pendulum. The characterization by the extended-bob pendulum ignores fewer of those features – it is more detailed than the one given by the simple pendulum but not thereby inconsistent with it. There is no need to privilege one characterization as a guide to what pendula are like rather than the other, since the characterizations are compatible. Nonetheless, the characterizations are both imperfect, in virtue of their being incomplete; and the extended-bob pendulum is less imperfect than the simple pendulum, in virtue of representing a feature of pendula not represented by the simple pendulum.

6.4 Conclusion

Most scientific hypotheses are idealized. When these hypotheses are explanatory, they have a claim to be guides to what the world is like. But not every explanatory idealized hypotheses can be a guide to ontology, because sometimes such hypotheses are incompatible with each other. The chapter argues that treating idealizations as abstractions rather than distortions provides an account of the connection between idealized hypotheses and ontology superior to the ones suggested by Sklar and Teller. Hence, insofar as idealized hypotheses can be guides to what the world is like, and insofar as inference to the best explanation is a cogent form of inference, this chapter provides a further motivation for interpreting idealizations as abstractions rather than as distortions.

CHAPTER 7

RESOLVING THE BAYESIAN PROBLEM OF IDEALIZATION

Although the preceding chapter provides some support for treating eliminable idealizations as abstractions rather than distortions, it is open to two objections. First, one might argue that the chapter only shows that idealizations involved in hypotheses that are candidate ontological guides are abstractions; and one might deny altogether that any idealized hypotheses are candidate ontological guides. Second, one might note that the argument from the previous chapter shows some eliminable idealizations to be abstractions only under the supposition that inference to the best explanation is cogent; denying the cogency of this inference, one might object that the argument fails to establish that any eliminable idealizations are abstractions.

Rather than rebut these potential objections, this chapter provides one further argument in favor of treating eliminable idealizations as abstractions rather than distortions, by showing that such an interpretation solves another problem concerning the role of idealization in science. This problem does not presume that some idealized hypotheses are candidate ontological guides; and the proposed solution does not depend upon the cogency of inference to the best explanation. The problem originates from the intuition that some idealized hypotheses can be confirmed at least to some degree. The widespread belief that scientific hypotheses can be confirmed, and the fact that most scientific hypotheses are idealized, lends a high degree of plausibility to this intuition. Moreover, since the problem to be addressed in this chapter lies within the province of confirmation theory, the proposed solution shows the usefulness of treating idealizations as abstractions for issues not concerned with scientific explanation.

The focus of this chapter is a problem due to Michael Shaffer, who challenges Bayesian confirmation theorists to show how at least some idealized hypotheses have at least some degree of confirmation ([105]). He argues that, in order to accomplish this task, one must either develop a coherent proposal for how to assign prior probabilities to counterfactual conditionals or abandon Bayesianism. This chapter develops a Bayesian reply to Shaffer's challenge that avoids the issue of how to assign prior probabilities to counterfactuals by treating idealized hypotheses as abstract descriptions and idealizations as abstractions. The reply allows Bayesians to assign non-zero degrees of confirmation to idealized hypotheses and to capture the intuition that less idealized hypotheses tend to be better confirmed than their more idealized counterparts.

7.1 The Bayesian Problem of Idealization

According to Bayesian confirmation theory (hereafter: Bayesianism), the posterior probability of an hypothesis H given evidence E, $\Pr(H \mid E)$, determines how well E confirms H. (Bayesians typically interpret the function $\Pr(-)$ as a subjective probability.) And one hypothesis $H1$ is better supported by evidence E than rival hypothesis $H2$ just if E confirms $H1$ more than $H2$. That said, there is no general agreement among Bayesians on how to measure the degree to which evidence supports an hypothesis. Some Bayesians favor a difference measure, according to which the degree that E confirms H is equal to $\Pr(H \mid E)$ - $\Pr(H)$. Others favor a normalized difference measure, according to which the degree that E confirms H is equal to $\Pr(H \mid E)$ - $\Pr(H \mid \text{not-}E)$. Some favor a ratio measure: $\Pr(H \mid E)$ / $\Pr(H)$. And some favor a likelihood measure: $\Pr(H \mid E)$*[1 - $\Pr(H)$] / [1 - $\Pr(H \mid E)$]*$\Pr(H)$.

Despite these differences, all Bayesian measures involve the quantity $\Pr(H \mid E)$. This is the fact that is relevant to what Shaffer calls the Bayesian problem of idealization. The problem, in a nutshell, is this: Bayesian confirmation theory seems to entail that the posterior probability of every idealized hypothesis is undefined; this entails that Bayesianism is unable to account for the fact that some idealized hypotheses can be confirmed at least to some degree; and this entails that Bayesianism is unable to make sense of the intuition that less idealized hypotheses tend to be better confirmed than their more idealized counterparts.[99]

These unpleasant consequences result from the assumption that every idealized hypothesis is a counterfactual conditional in which the antecedent is a set of idealizing conditions – such as "Each particle's radius $r \to 0$" – and the consequent is a set of claims that are true under those conditions – such as the ideal gas equation PV = NkT. (More on this assumption in the next section.) If idealized hypotheses have the form $A > C$ (where '>' is the symbol for counterfactual conditionals), then the posterior probability of an idealized hypothesis relative to evidence E has the

[99]Shaffer further argues that since most scientific hypotheses are idealized, Bayesianism entails that "few, if any, scientific theories have ever been confirmed to any extent whatsoever" (p. 45). This corollary makes Shaffer's argument more interesting. But I ignore it as incidental to the prior issue of whether Bayesianism can accommodate the fact that at least some idealized hypotheses are confirmed to at least some degree.

form $\Pr(A > C \mid E)$. According to Bayes' Theorem, the posterior probability $\Pr(A > C \mid E)$ is equal to $\Pr(E \mid A > C)$*$\Pr(A > C)$ / $\Pr(E)$. The problem for Bayesians is that there is no extant, coherent suggestion for how to assign prior probabilities to counterfactuals – that is, for how to assign values to $\Pr(A > C)$.[100] Since the posterior probability $\Pr(A > C \mid E)$ is defined only if the prior probability $\Pr(A > C)$ is defined, it seems that the posterior probabilities of idealized hypotheses are undefined. And, at least for Bayesians, this entails that the degree to which any given idealized hypothesis is confirmed by evidence is undefined. Given that most scientific hypotheses are idealized in some way, Bayesianism seems to entail that most scientific hypotheses cannot be confirmed.

Bayesians thus confront an apparent trilemma: either develop a coherent proposal for how to assign prior probabilities to counterfactuals; or embrace the counterintuitive result that idealized hypotheses cannot be confirmed; or reject Bayesianism. There is also a fourth option, developed in the remainder of this paper: reject the assumption that idealized hypotheses are counterfactual conditionals.

7.2 Motivating the Appeal to Counterfactuals

According to Shaffer, "When we claim that a theory holds in some idealized model, or under some idealizing conditions, we are claiming that a theory is true only on the basis of one or more counterfactual simplifying assumptions or conditions" (p. 41). He insists that "theories incorporating idealizing conditions ought to be construed as counterfactuals" (p. 43), adding that "this thesis is not open to question". Among

[100]Shaffer considers and rejects three suggestions. I endorse those considerations. Shaffer does not consider treating idealized hypotheses as counterdoxastic conditionals. Since I am unaware of anyone who has developed this proposal, I leave an evaluation of the proposal as a project for some other time.

others, Frederick Suppe ([115]) and Ilkka Niiniluoto ([87]) also claim that idealized hypotheses are counterfactuals. None of these authors motivates this claim. Perhaps it is worth pausing to fill the lacuna. The motivation seems to come in two stages: the first stage motivates treating idealized hypotheses as conditionals; the second stage motivates treating these conditionals as counterfactuals.

Stage One. Claims obtained through appeal to idealizations tend to be false of real systems (despite sometimes being "close enough" to the truth for various purposes). The equation for the simple pendulum is false of most real pendula, the ideal gas equation is false of most real gases, etc. If idealized hypotheses are the claims obtained through appeal to idealizations rather than conditionals in which such claims are the consequents, then most idealized hypotheses have a null posterior probability, in virtue of being inconsistent with available evidence. This result violates the intuitions that less idealized hypotheses tend to be better confirmed than their more idealized counterparts and that some idealized hypotheses have non-zero posterior probabilities relative to available evidence. Treating idealized hypotheses as conditionals rather than the claims obtained through appeal to idealizations allows them to be consistent with available evidence despite their consequents being inconsistent with that evidence.

Stage Two. If idealized hypotheses are conditionals (in which the antecedent is a set of idealizing conditions), then it is better to treat such conditionals as counterfactual rather than material. Since idealizing conditions typically are taken to be false, idealized hypotheses would be trivially true in virtue of the falsity of their antecedents if such hypotheses were material conditionals.[101] This result is unsatisfactory insofar as intuitions suggest that such hypotheses might be false. Moreover, saying that all idealized hypotheses are vacuously true in virtue of having false antecedents entails that the posterior probability of every idealized hypothesis is unity relative to any evidence whatsoever. For most measures of degree of confirmation, this means that differences in the degree to which competing idealized hypotheses are confirmed depends entirely upon the prior probabilities of those hypotheses – a counterintuitive result insofar as one expects evidence to play at least some role in determining rankings of the degrees to which competing hypotheses are confirmed.

7.3 Solving the Problem

Until Bayesians develop a coherent proposal for how to assign prior probabilities to counterfactuals, and unless Bayesians want to deny that at least some idealized hypotheses can be confirmed, they should reject the treatment of such hypotheses as counterfactual conditionals. So, for example, they should reject treating the ideal gas law as an hypothesis of the form "If such and such idealizing conditions were to obtain, then the ideal gas equation would be true". They should reject treating the law of motion for simple pendula as an hypothesis of the form "If such and such idealizing conditions were to obtain, then the equation of motion for the simple pendulum would be true". And so on.

At the same time, a Bayesian treatment of idealized hypotheses should not run afoul of the considerations raised in the previous section. A satisfactory Bayesian

[101] Leszek Nowak ([88], p. 136) recommends that, faced with the potential of idealized hypotheses being trivially true, we should continue to treat idealized hypotheses as material conditionals and revise the usual definition of truth.

treatment should not entail that most idealized hypotheses have a null posterior probability. Nor should it entail that they are vacuously true. (This latter desideratum seems to require not treating idealized hypotheses as conditionals, in which case the ideal gas law just is the ideal gas equation and the law of motion for simple pendula just is the equation of motion for the simple pendulum.)

All of these constraints can be met by treating idealized hypotheses as abstract descriptions and idealizations as abstractions. This solution abandons the problem of how to assign prior probabilities to counterfactual conditionals, because it rejects the assumption that idealized hypotheses are counterfactuals.

If idealizations are abstractions, then idealized hypotheses are not conditionals in which the antecedent is a set of idealizing conditions. For the antecedent of a conditional must be a set of statements; but idealizations are not statements if they are abstractions. Instead, they are more like "inference tickets" that transform one description of a system into a (more idealized) description that ignores certain features of the system. So treating idealizations as abstractions allows Bayesians to avoid worries about how to assign prior probabilities to counterfactual conditionals.

Moreover, if idealized hypotheses are abstract descriptions, there is nothing mysterious about how to determine their posterior probabilities via Bayes' Theorem. Prior probabilities are to be assigned to idealized hypotheses in the same way that such probabilities are assigned to incomplete or partial descriptions. And the conditional probability of the evidence given an idealized hypothesis (abstract description) will depend upon whether the features ignored by the hypothesis are relevant to the evidence (because the truth of an abstract description depends upon the relevance of the details it ignores). For instance, if the evidence shows only that there is a rough, qualitative proportionality between a specific real pendulum's period and the distance between its pivot and center of mass, then the conditional probability of this evidence given the law of motion for the simple pendulum is unity, since that law entails such a proportionality. But if the evidence also includes data about the exact period of a particular real pendulum, the conditional probability of this evidence given the law of motion for the simple pendulum is probably null, since most likely the equation for the simple pendulum predicts an incorrect period: it is probably false with respect to the exact period of the real pendulum in virtue of ignoring features that are relevant to the exact period.

Bayesians can avoid assigning zero as the conditional probability of the evidence given an idealized hypotheses by calculating this probability relative to a subset of all available evidence, such as evidence for which the features ignored by the idealized hypothesis are irrelevant. This selective attention to the evidence seems to accord with scientific practice. For instance, there is good reason to think that the current best scientific theories – general relativity and quantum field theory – are idealized. General relativity ignores quantum effects with the idealization that Planck's constant $h \to 0$, so that the Compton wavelength $\lambda_C = h/mc \to 0$. (The Compton wavelength is roughly the distance scale at which quantum field theory becomes important for understanding the behavior of objects with mass m.) Quantum field theory ignores gravitational effects with the idealization that Newton's gravitational constant $G \to 0$, so that the Schwarzschild radius $r_S = 2GM/c^2 \to 0$. (The Schwarzschild radius is roughly the distance scale at which general relativity becomes important for understanding the behavior of objects with mass M.) But the failures of general relativity to accommodate quantum effects and of quantum field theory to accommodate gravitational effects are not taken to disconfirm those theories. Instead, the range of each theory is restricted to phenomena for which quantum or gravitational effects are irrelevant, respectively. And this restriction permits a restriction of the range of phenomena – or sources of evidence – that are eligible for confirming or disconfirming each theory. Similar restrictions occur with effective field theories, such as the Euler-Heisenberg theory for photon-photon scattering. This theory's range is restricted to phenomena in which the electron field is irrelevant to photon interactions (i.e., phenomena that occur at energy scales below the threshold for electron production), so that the theory is not disconfirmed by phenomena in which the electron field is relevant to photon interactions.

Finally, if idealized hypotheses are abstract descriptions, it is possible to make sense of the intuition that less idealized hypotheses tend to be better confirmed than their more idealized counterparts. Consider the law of motion for the (undamped) simple pendulum and the law of motion for the damped pendulum. The latter is less idealized than the former, in virtue of taking into account the amount of damping on pendula. So it is to be expected that set of evidence for which the features ignored by the law of motion for the damped pendulum are irrelevant is larger than the set of evidence for which the features ignored by the law of motion for the simple pendulum are irrelevant: the former set contains the latter plus evidence about phenomena in which damping is relevant. Relative to this larger set of evidence, Bayesians can expect the law of motion for the damped pendulum to be better supported than the law of motion for the simple pendulum, since the larger set of evidence is probably inconsistent with the predictions obtained from the law of motion for the simple pendulum.

CHAPTER 8

CONCLUDING REMARKS

The aim of this dissertation is to contribute to the philosophical literature on explanation and the role of idealization. To this end, the dissertation focuses on two putative explanations from statistical mechanics: the standard account of phase transitions and the Boltzmannian account of irreversible behavior. Like many explanatory accounts in physics, these accounts are idealized: the descriptions they invoke are not entirely true of the systems in which their explananda occur. Unlike most idealized explanations, however, certain idealizations that occur in these accounts are ineliminable: the only way to obtain a description (let alone an explanation) of phase transitions and irreversibility, according to the accounts, is to invoke idealizing assumptions – namely, the thermodynamic limit and the Boltzmann-Grad limit, respectively.

Ineliminably idealized explanations are not well-understood from a philosophical point of view. Several philosophical accounts of idealized explanation fail to accommodate them. The dominant understanding of idealizations as falsehoods precludes their existence. Moreover, there are powerful arguments, based upon Earman's Principle and Constructionism, against the possibility of phenomena that require for their description appeal to idealization. And there are error-theoretic approaches that suggest the statistical mechanical accounts of phase transitions and irreversibility mischaracterize those phenomena by taking mathematical precision too seriously.

Despite the foregoing philosophical difficulties, the explanatory nature of the accounts of phase transitions and irreversibility can be understood by taking their characteristic idealizations to be abstractions and treating idealized explanations as a kind of incomplete explanation. According to this approach, idealizations like the thermodynamic limit are not false claims about the way the world really is. Rather, they are devices for ignoring certain details about the world. This non-standard interpretation of idealizations is consistent with the mathematical role served by the thermodynamic and Boltzmann-Grad limits in their respective accounts. And it provides a response to those who deny the existence of phase transitions and genuinely irreversible behavior on the grounds that their descriptions require, in principle, an appeal to idealizations. An interesting corollary of this approach is that sometimes the only way to describe a phenomenon correctly is to describe it incompletely – sometimes microscopic details obscure what is to be explained.

Furthermore, the treatment of idealizations as abstractions rather than distortions solves two independent problems regarding the role of idealization in science. The first is the problem of how to select a guide to what the world is like from competing but incompatible idealized hypotheses. If multiple competing hypotheses explanatorily characterize the same system, at most one can be a guide to the nature of that system. If idealizations are false, inference to the best explanation is not a cogent method for privileging one explanatory idealized hypothesis over its competitors as a guide to what the world is like; and extant proposals about when explanatory idealized hypotheses are guides to what the world is like – proposals that do not appeal to explanatory considerations in privileging one hypothesis over its competitors – have problems of their own. But if idealizations are abstractions, then inference to the best explanation can be cogent when the hypotheses involved are idealized, because idealized hypotheses need not be false if idealizations omit details without falsifying them.

The second problem solved by treating idealizations as abstractions rather than distortions is the Bayesian problem of idealization. Scientific hypotheses are capable of being confirmed to at least some degree; and most scientific hypotheses are idealized. If idealizations are distortions, however, it does not appear possible to assign any posterior probability to any idealized hypothesis; and so it does not seem possible for idealized hypotheses to be confirmed. If, in contrast, idealizations are abstractions, then posterior probabilities can be assigned to idealized hypotheses in the same way they are assigned to incomplete descriptions, thereby allowing non-zero degrees of confirmation to be assigned to such hypotheses.

Steven Orszack suggests two questions relevant to assessing any methodological thesis about science ([90], p. 479):

1. Is the claim conceptually coherent and empirically adequate?

2. Does the claim lead to normative procedures that improve the quality of science and lead to a better understanding of nature?

A central methodological claim of this dissertation is that some scientific explanations are ineliminably idealized. Granting that this claim has been shown to be coherent and true, it has several normative consequences. First, certain research programs are not defective on a priori grounds, merely in virtue of involving an ineliminable appeal to certain idealizations – and hence these programs need not be abandoned. Second, one need not abandon mathematical precision to characterize the explananda for such explanations. Third, the defense of this claim against objections provides a rationale for why, sometimes, the explananda of such explanations are called emergent; and the defense gives some sense to this way of speaking.

BIBLIOGRAPHY

[1] Ralph H. Abraham and Christopher D. Shaw. *Dynamics: The Geometry of Behavior*. Addison-Wesley, New York, 2nd edition, 1992. 41

[2] P. Achinstein and O. Hannaway, editors. *Observation, Experiment and Hypothesis in Modern Physical Science*. MIT Press, Cambridge, MA, 1985. 254

[3] P.W. Anderson. More Is Different. *Science*, 177:292–296, 1972. 200

[4] Robert W. Batterman. Explanatory Instability. *Nous*, 26(3):325–348, 1992. 17

[5] Robert W. Batterman. Asymptotics and the Role of Minimal Models. *British Journal for Philosophy of Science*, 53:21–38, 2002. 143, 175

[6] Robert W. Batterman. *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence*. Oxford University Press, New York, 2002. 17

[7] Robert W. Batterman. Critical Phenomena and Breaking Drops: Infinite Idealizations in Physics. *Studies in History and Philosophy of Modern Physics*, 36:225–244, 2005. 190

[8] Robert W. Batterman. 'Fundamental' Theory: Molecular Dynamics vs. Hydrodynamics. *Erkenntnis*, forthcoming. 187

[9] Alexander Bird. Explanation and Laws. *Synthese*, 120:1–18, 1999. 136

[10] R.A. Blythe and M.R. Evans. The Lee-Yang Theory of Equilibrium and Nonequilibrium Phase Transitions. *Brazilian Journal of Physics*, 33(3):464–475, 2003. 71

[11] Paul A. Boghossian. The Status of Content. *The Philosophical Review*, 99(2):157–184, 1990. 183, 184

[12] Gerald L. Bradley and Karl J. Smith. *Calculus*. Prentice-Hall, Upper Saddle River, NJ, 2nd edition, 1999. 41

[13] May Brodbeck. Introduction: Theory Construction. In *Readings in the Philosophy of Science* [14]. 126

[14] May Brodbeck. *Readings in the Philosophy of Science*. The MacMillan Company, New York, 1968. 251

[15] Craig Callender. Reducing Thermodynamics to Statistical Mechanics: The Case of Entropy. *The Journal of Philosophy*, 96(7):348–373, 1999. 90

[16] Craig Callender. Taking Thermodynamics (Too) Seriously. *Studies in the History and Philosophy of Modern Physics*, 32(4):539–553, 2001. 178, 179, 180, 189

[17] Nancy Cartwright. *Nature's Capacities and their Measurement*. Clarendon Press, Oxford, 1989. 9, 134

[18] Nancy Cartwright. *The Dappled World: A Study of the Boundaries of Science*. Cambridge University Press, New York, 1999. 134

[19] Nancy Cartwright. Models and the limits of theory: quantum Hamiltonians and the BCS model of superconductivity. In Morgan and Morrison [84], pages 241–281. 134

[20] Carlo Cercignani. *The Boltzmann Equation and Its Application*. Springer-Verlag, New York, 1988. 72, 97, 101, 165, 168

[21] Carlo Cercignani, V.I. Gerasimenko, and D. Ya. Petrina. *Many-Particle Dynamics and Kinetic Equations*. Kluwer Academic Publishers, Boston, 1997. 98, 104, 109

[22] Carlo Cercignani, Reinhard Illner, and Mario Pulvirenti. *The Mathematical Theory of Dilute Gases*. Springer-Verlag, New York, 1994. 98, 109

[23] Anjan Chakravartty. The Semantic or Model-Theoretic View of Theories and Scientific Realism. *Synthese*, 127(3), 2001. 128

[24] Debashish Chowdhury and Dietrich Stauffer. *Principles of Equilibrium Statistical Mechanics*. Wiley-VCH, New York, 2000. 65

[25] Barry A. Cipra. An Introduction to the Ising Model. *American Mathematical Monthly*, 94(10):937–959, 1987. 71

[26] G.A.T.F. da Costa and A.L. Maciel. Combinatorial Formulation of Ising Model Revisited. *Revista Brasileira de Ensino de Fisica*, 25(1):49–61, 2003. 80

[27] Kevin Davey. Is Mathematical Rigor Necessary in Physics? *British Journal for the Philosophy of Science*, 54(3):439–463, 2003. 183

[28] P.C.W. Davies. *The Physics of Time Asymmetry.* University of California Press, Berkeley, 1977. 181

[29] J.R. Dorfman. *An Introduction to Chaos in Nonequilibrium Statistical Mechanics.* Cambridge University Press, New York, 1999. 110

[30] John Earman. Laws, Symmetry, and Symmetry Breaking; Invariance, Conservation Principles, and Objectivity. http://philsci-archive.pitt.edu/archive/00000878/, 2002. 194

[31] R. Esposito, J.L. Lebowitz, and R. Marra. On the Derivation of Hydrodynamics from the Boltzmann equation. *Physics of Fluids*, 11(8):2354–2366, 1999. 97

[32] Herbert Feigl and May Brodbeck, editors. *Readings in the Philosophy of Science.* Appleton-Century-Crofts, New York, 1953. 253

[33] M. Ferrari and I. O.Stamatescu, editors. *Symbol and Physical Knowledge: On the Conceptual Structure of Physics.* Springer, New York, 2002. 253

[34] Peter French, editor. *Midwest Studies in Philosophy, Volume XVIII, Philosophy of Science.* University of Notre Dame Press, South Bend, IN, 1993. 253

[35] Michael Friedman. Explanation and Scientific Understanding. *The Journal of Philosophy*, 71(1):5–19, 1974. 171

[36] Giovanni Gallavotti. *Statistical Mechanics: A Short Treatise.* Springer, New York, 1999. 191

[37] Axel Gelfert. Mathematical Rigor in Physics: Putting Exact Results in Their Place. http://philsci-archive.pitt.edu/archive/00002160/, 2005. 74, 120, 121

[38] Ronald N. Giere. *Explaining Science: A Cognitive Approach.* University of Chicago Press, Chicago, 1988. 53

[39] Adolf Grünbaum and Wesley C. Salmon, editors. *The Limitations of Deductivism.* University of California Press, Berkeley, 1988. 253

[40] Stewart Harris. *An Introduction to the Theory of the Boltzmann Equation.* Holt, Rinehart and Winston, Inc., New York, 1971. 98

[41] Stephan Hartmann. Effective Field Theories, Reductionism and Scientific Explanation. *Studies in History and Philosophy of Modern Physics*, 32(2):264–304, 2001. 141, 143

[42] Daniel Heard. A New Problem for Ontological Emergence. *The Philosophical Quarterly*, 56(222):55–62, 2006. 203, 206

[43] Carl G. Hempel. *Aspects of Scientific Explanation and other Essays in the Philosophy of Science*. Free Press, New York, 1965. 6, 36, 140

[44] Carl G. Hempel. Provisos: A Problem Concerning the Inferential Function of Scientific Theories. In Grünbaum and Salmon [39], pages 19–36. 11

[45] Carl G. Hempel and Paul Oppenheim. The Logic of Explanation. In Feigl and Brodbeck [32], pages 319–352. 6

[46] Terrell L. Hill. *Thermodynamics of Small Systems, Part I*. W.A. Benjamin, Inc., New York, 1963. 156

[47] R.I.G. Hughes. Theoretical Explanation. In French [34], pages 132–153. 51, 171

[48] Paul W. Humphreys. Scientific Explanation: The Causes, Some of the Causes, and Nothing But the Causes. In Kitcher and Salmon [58], pages 283–305. 13

[49] Andreas Hüttemann. Idealizations in Physics. In Ferrari and O.Stamatescu [33], pages 177–192. 3

[50] Martin R. Jones. Idealization and Abstraction: A Framework. In *Idealization XII: Correcting the Model* [51], pages 173–217. 26

[51] Martin R. Jones and Nancy Cartwright. *Idealization XII: Correcting the Model*. Rodopi, New York, 2005. 253

[52] Xiao jun Wang, Chris Schmidtt, and Marvin Payne. Oscillations with Three Damping Effects. *European Journal of Physics*, 23:155–164, 2002. 39

[53] Leo P. Kadanoff. *Statistical Physics: Statics, Dynamics, and Renormalization*. World Scientific, Singapore, 2000. 71

[54] A.C. Kakas and F. Sadri, editors. *Computational Logic: From Logic Programming into the Future*. Springer-Verlag, Berlin, 2002. 256

[55] Jaegwon Kim. Making Sense of Emergence. *Philosophical Studies*, 95:3–36, 1999. 202, 205, 206

[56] Philip Kitcher. Explanatory Unification. *Philosophy of Science*, 48:507–531, 1981. 7, 171

[57] Philip Kitcher. Explanatory Unification and the Causal Structure of the World. In Kitcher and Salmon [58], pages 410–505. 7, 8, 47

[58] Philip Kitcher and Wesley C. Salmon, editors. *Minnesota Studies in the Philosophy of Science, Volume XIII: Scientific Explanation*. University of Minnesota Press, Minneapolis, 1989. 253

[59] Jeffrey Koperski. Models, Confirmation, and Chaos. *Philosophy of Science*, 65:624–648, 1998. 36

[60] Oscar E. Lanford, III. On a Derivation of the Boltzmann Equation. In Lebowitz and Montroll [66], pages 1–17. 98

[61] Ronald Laymon. Idealization, Explanation, and Confirmation. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Volume One: Contributed Papers:336–350, 1980. 36, 38

[62] Ronald Laymon. Idealization and the Testing of Theories by Experimentation. In Achinstein and Hannaway [2], pages 147–173. 36, 38

[63] Ronald Laymon. Applying Idealized Scientific Theories to Engineering. *Synthese*, 81:353–371, 1989a. 36

[64] Ronald Laymon. Cartwright and the Lying Laws of Physics. *The Journal of Philosophy*, 86(7):353–372, 1989b. 36, 37

[65] Ronald Laymon. Experiment and the Legitimacy of Idealization. *Philosophical Studies*, 77:353–375, 1995. 9

[66] J.L. Lebowitz and E.W. Montroll, editors. *Nonequilibrium Phenomena I: The Boltzmann Equation*. North Holland, Amsterdam, 1983. 254

[67] Joel L. Lebowitz. Statistical Mechanics: A Selective Review of Two Central Issues. *Reviews of Modern Physics*, 71:2:S346–S357, 1999. 90, 157

[68] T.D. Lee. and C.N. Yang. Statistical Theory of Equations of State and Phase Transitions. II. Lattice Gas and Ising Model. *Physical Review*, 87(3):410–419, 1952. 71

[69] David Lewis. *Philosophical Papers, Volume II*. Oxford University Press, New York, 1986a. 254

[70] David Lewis. Causal Explanation. In *Philosophical Papers, Volume II* [69], pages 214–240. 7

[71] David Lewis. *Counterfactuals*. Blackwell Publishers, Malden, MA, 2001. 54

[72] Chuang Liu. Approximation, Idealization, and Laws of Nature. *Synthese*, 118(2):229–256, 1999. 189

[73] Chuang Liu. Infinite Systems in SM Explanations: Thermodynamic Limit, Renormalization (semi-) Groups, and Irreversibility. *Philosophy of Science Proceedings*, 68:S325–S344, 2001. 178, 182

[74] Chuang Liu and Gérard G. Emch. Explaining Quantum Spontaneous Symmetry Breaking. *Studies in History and Philosophy of Modern Physics*, 36(1):137–163, 2005. 188, 195, 196

[75] P.L. Lyons and N. Masmoudi. From the Boltzmann Equations to the Equations of Incompressible Fluid Mechanics, I. *Arch. Rational Mech. Anal.*, 158:173–193, 2001. 97

[76] R.B. Marcus, editor. *Logic, Methodology, and Philosophy of Science*. North-Holland, Amsterdam, 1986. 256

[77] Michael R. Matthews. *Time for Science Education: How Teaching the History and Philosophy of Pendulum Motion can Contribute to Science Literacy*. Kluwer Academic / Plenum Publishers, New York, 2000. 19

[78] J.E. Mayer. *Equilibrium Statistical Mechanics*. Pergamon Press, New York, 1968. 180

[79] Ernan McMullin. Galilean Idealization. *Studies in the History and Philosophy of Science*, 16:247–273, 1985. 126

[80] Paul E. Meehl and Wilfrid Sellars. The Concept of Emergence. pages 239–252. 1956. 208

[81] Alexander Miller. Wright's Argument Against Error-Theories. *Analysis*, 62(2):98–103, 2002. 186

[82] R.A. Minlos. *Introduction to Mathematical Statistical Physics*. American Mathematical Society, Providence, RI, 2000. 182

[83] Sandra D. Mitchell. Integrative Pluralism. *Biology and Philosophy*, 17:55–70, 2002. 224

[84] Mary S. Morgan and Margaret Morrison, editors. *Models as Mediators: Perspectives on Natural and Social Science*. Cambridge University Press, New York, 1999. 251, 255

[85] Margaret Morrison. Models as Autonomous Agents. In Morgan and Morrison [84], pages 38–65. 9, 129

[86] F. A. Muller. The Deep Black Sea: Observability and Modality Afloat. *British Journal for Philosophy of Science*, 56:61–99, 2005. 54

[87] Ilkka Niiniluoto. Theories, Approximations, and Idealizations. In Marcus [76], pages 255–289. 240

[88] Leszek Nowak. *The Structure of Idealization: Towards a Systematic Interpretation of the Marxian Idea of Science*. Reidel, Dordrecht, 1980. 24, 26, 241

[89] Onora O'Neill. *Bounds of Justice*. Cambridge University Press, New York, 2000. 126

[90] Steven Orzack. Discussion: What, if anything, is 'the strategy of model building in population biology'? a comment on Levinas (1996) and Odenbaugh (2003). *Philosophy of Science*, 72(3):479–485, 2005. 248

[91] Stephen Pepper. Emergence. *Journal of Philosophy*, 23:241–245, 1926. 206, 207, 208

[92] Michael H. Peters. *Molecular Thermodynamics and Transport Phenomena: Complexities of Scales in Space and Time*. McGraw-Hill, New York, 2005. 90, 95

[93] Karl R. Popper. *The Poverty of Historicism*. Routledge and Kegan Paul, London, 1957. 24

[94] Stathis Psillos. On Van Fraassen's Critique of Abductive Reasoning. *The Philosophical Quarterly*, 46:31–47, 1996. 215

[95] Stathis Psillos. *Causation and Explanation*. McGill-Queen's University Press, Ithaca, 2002. 6

[96] Stathis Psillos. Simply the Best: A Case for Abduction. In Kakas and Sadri [54], pages 605–625. 230

[97] Peter Railton. Probability, Explanation, and Information. *Synthese*, 48:233–256, 1981. 36, 150

[98] Bimalendu N. Roy. *Fundamentals of Classical and Statistical Thermodynamics*. John Wiley and Sons, New York, 2002. 187

[99] Alexander Rueger. Robust Supervenience and Emergence. *Philosophy of Science*, 67:466–489, 2000. 86

[100] Alexander Rueger. Reduction, Autonomy, and Causal Exclusion among Physical Properties. *Synthese*, 139:1–21, 2004. 209, 211

[101] Alexander Rueger and W. David Sharp. Idealization and Stability: A Perspective from Nonlinear Dynamics. In Shanks [106], pages 201–216. 40, 42, 43, 45

[102] Laure Saint-Raymond. From the BGK Model to the Navier-Stokes Equations. *Annales Scientifiques De L'École Normale Supérieure*, 4(36):271–317, 2003. 97

[103] Wesley C. Salmon. *Causality and Explanation*. Oxford University Press, New York, 1998. 7, 151

[104] Daniel V. Schroeder. *An Introduction to Thermal Physics*. Addison Wesley Longman, New York, 2000. 69

[105] Michael Shaffer. Bayesian Confirmation of Theories that Incorporate Idealizations. *Philosophy of Science*, 68:36–52, 2001. 215, 237

[106] Niall Shanks, editor. *Idealization IX: Idealization in Contemporary Physics*. Rodopi, Amsterdam-Atlanta, GA, 1998. 256

[107] Michael Silberstein and John McGeever. The Search for Ontological Emergence. *The Philosophical Quarterly*, 49:182–200, 1999. 202

[108] Lawrence Sklar. *Physics and Chance: Philosophical Issues in the Foundations of Statistical Mechanics*. Cambridge University Press, New York, 1993. 89, 99

[109] Lawrence Sklar. *Theory and Truth: Philosophical Critique within Foundational Science*. Oxford University Press, New York, 2000. 54, 219

[110] Lawrence Sklar. Dappled Theories in a Uniform World. *Philosophy of Science*, 70:424–441, 2003. 214, 219, 229

[111] Elliot Sober. *Simplicity*. Oxford University Press, Oxford, 1975. 3

[112] Elliot Sober. Physicalism from a Probabilistic Point of View. *Philosophical Studies*, 95:135–174, 1999. 202

[113] Michael Strevens. The Causal and Unification Approaches to Explanation Unified – Causally. *Nous*, 38(1):154–176, 2004. 146

[114] Daniel F. Styer. What good is the thermodynamic limit? *American Journal of Physics*, 72(1):25–29, 2004. 75

[115] Frederick Suppe. *The Semantic Conception of Theories and Scientific Realism*. University of Illinois Press, Chicago, 1989. 240

[116] Paul Teller. Twilight of the Perfect Model Model. *Erkenntnis*, 55:393–415, 2001. 52, 222

[117] Paul Teller. How We Dapple the World. *Philosophy of Science*, 71:425–447, 2004. 220, 221, 222, 224, 225

[118] Neil Tennant. Perfect Validity, Entailment, and Paraconsistency. *Studia Logica*, XLIII:179–198, 1984. 146

[119] Paul Thagard. The Best Explanation: Criteria for Theory Choice. *Journal of Philosophy*, 75(2):76–92, 1978. 213

[120] Paul Thagard. *Coherence in Thought and Action*. MIT Press, Cambridge, MA, 2000. 231

[121] Colin J. Thompson. *Mathematical Statistical Mechanics*. The Macmillan Company, New York, 1972. 105, 110

[122] Bas C. van Fraassen. *Laws and Symmetry*. Oxford University Press, New York, 1989. 6, 215

[123] James Woodward. *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press, New York, 2003. 171

[124] C.N. Yang and T.D. Lee. Statistical Theory of Equations of State and Phase Transitions. I. Theory of Condensation. *Physical Review*, 87(3):404–409, 1952. 71

[125] Sang Wook Yi. Reduction of Thermodynamics: A Few Problems. *Philosophy of Science*, 70:1028–1038, 2003. 178, 181