ASSUME A SPHERICAL COW:
STUDIES ON REPRESENTATION AND IDEALIZATION

by

Elay Shech

B.A., Philosophy, Boston University, 2007

B.A., Physics, Boston University, 2007

M.A., Philosophy, University of Pittsburgh, 2012

M.A., History and Philosophy of Science, University of Pittsburgh, 2012

Submitted to the Graduate Faculty of the

Kenneth P. Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy in History and Philosophy of Science

University of Pittsburgh

2015

UNIVERSITY OF PITTSBURGH

KENNETH P. DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Elay Shech

It was defended on

March 30th, 2015

and approved by

John D. Norton, PhD, History and Philosophy of Science

John Earman, PhD, History and Philosophy of Science

Robert Batterman, PhD, Philosophy

Giovanni Valente, PhD, Philosophy

Laura Ruetsche, PhD, Philosophy (University of Michigan)

Dissertation Advisors: John D. Norton, PhD, History and Philosophy of Science

John Earman, PhD, History and Philosophy of Science

Copyright © by Elay Shech

2015

ASSUME A SPHERICAL COW:

STUDIES ON REPRESENTATION AND IDEALIZATION

Elay Shech, PhD

University of Pittsburgh, 2015

This dissertation concerns the philosophical underpinnings of representation and idealization in science. I begin by looking at the philosophical debate revolving around phase transitions and use it as a foil to bring out what I take to be most interesting about phase transitions, namely, the manner by which they illustrate *the problem of essential idealizations*. I continue to solve the problem in several steps. First, I conduct an interdisciplinary comparative study of different types of representations (e.g., mental, linguistic, pictorial) and consequently promote a content-based account of scientific representation intended to accommodate the practice of idealization and misrepresentation. I then critically asses the literature on idealizations in science in order to identify the manner by which to justify appeals to idealizations in science, and implement such techniques in two case studies that merit special attention: the Aharonov-Bohm effect and the quantum Hall effects. I proceed to offer a characterization of *essential idealizations* meant to alleviate the woes associated with said problem, and argue that particular types of idealizations, dubbed *pathological idealizations*, ought to be dispensed with. My motto is that idealizations are essential to explanation and representation, as well as to methodology and pedagogy, but they essentially misrepresent. Implications for the debate on platonism about mathematical objects are outlined.

Keywords: philosophy of physics, philosophy of science, philosophy of mathematics, representation, idealization, phase transitions, Aharonov-Bohm effect, fractional quantum statistics, quantum Hall effects.

Dedicated to my parents, Orit & Aharon Shech

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

This dissertation began as a seminar paper that I wrote for Jeremy Butterfield, while reading Robert Batterman's *The Devil in the Details*, and through discussions with John Earman and John Norton on the idealizations appealed to in the explanation of fractional quantum statistics and phase transitions, respectively. It evolved into a prospectus through a directed study conducted with Robert Batterman and Giovanni Valente, matured into its current state through the guidance of John Earman and John Norton, and has been especially influenced by the numerous works of Robert Batterman on idealization and explanation, John Earman's "Understanding Permutation Invariance in Quantum Mechanics," John Norton's "Approximation and Idealizations: Why the Difference Matters," and Laura Ruetsche's *Interpreting Quantum Theories*.

I spent eight years in University of Pittsburgh's History and Philosophy of Science doctoral program, the pinnacle of which is represented by this dissertation document. My time at HPS is marked with many marvelous and happy moments of illuminating conversations with the plethora of wonderful philosophers at Pitt, insightful reflections during long walks through Hillman Library, the Cathedral of Learning, and Frick Park, and an overall sense of deep and satisfying intellectual development. However, I have also had to face some of my most difficult, challenging, and frustrating moments to date. In this regard, there are several people who played an essential and indispensable role in helping me along the way, and who I wish to thank.

I have been especially lucky to have had the support, and the opportunity to interact with and learn from, some wonderful colleagues both in the realm of faculty members and graduate students. On the faculty side, I particularly wish to thank Gordon Belot, Jeremy Butterfield, Peter Machamer, Edouard Machery, Jim Lennox, and Paolo Palmieri for their support throughout the years and at different stages of the program. On the graduate student side, I am grateful to David de Bruijn, Peter Distelzweig, Balázs Gyenis, Eric Hatleback, Marina Baldissera Pacchetti, Tom Pashby, Bryan Roberts and Mike Tamir for their friendship and help. More generally, thank you to the entire generation of students composing the community, which I am proud to have been part of, manifested to a great deal in our weekly "Work-in-Progress" workshops. For me, an important part of surviving through graduate school included walking in the footsteps of those students who came before me, and here my greatest debt is to Eric Hatleback and Bryan Roberts, who helped guide me at each step of the process. Thanks also to Eric Angner, Robert Batterman, David de Bruijn, John Earman, Samuel Fletcher, Greg Gadenberger, Nicholaos Jones, Michael Miller, John Norton, Aaron Novick, Isabel Ranner, Bryan Roberts, Laura Ruetsche, Naharin Shech, Giovanni Valente, Mark Wilson, and James Woodward specifically for assistance with various aspects of this dissertation and many constructive comments on earlier versions.

Most of all, on the academic and intellectual front, I wish to thank my dissertation committee—Robert Batterman, John Earman, Laura Ruetsche, John Norton, and Giovanni Valente—for their unwavering help, support, and mentorship, in the form of countless conversations and correspondences, directed studies, and engaging seminars throughout the years. I am especially grateful to my co-advisers John Earman and John Norton for teaching me how to do philosophy of science, working with me through the finer details, and encouraging me to pursue the questions that I found most fascinating. Of special note is their untiring patience. It is difficult to count the amount of times that I strode into John Norton's office with a bleak mood and what seemed like an insurmountable challenge, and always left with renewed inspiration and determination. Articulating the extent to which I am grateful to "the Johns" would necessitate adding additional chapters to this dissertation. Instead, I will attempt to sum it all up by noting that, when I look back at the last eight years, the single greatest experience, for which I am most thankful, honored, and proud, is that I have had the opportunity to be a student of John Earman and John Norton.

Finally, I wish to thank my family: my mother and father, Orit and Aharon Shech, for the opportunities they have provided, and together with my sisters and girlfriend, Naharin Shech, Nadine Shech, and Isabel Ranner, for their steadfast emotional support, constant encouragement, and unconditional love. This dissertation would not have been possible without your support, and for that I am eternally grateful.

1.0 INTRODUCTION

Milk production at a dairy farm was so low that the farmer wrote to the local university, asking help from scientists. A multidisciplinary team of professors was assembled, headed by a theoretical physicist, and two weeks of intensive on-site investigation took place. The scholars then returned to the university, notebooks crammed with data, where the task of writing the report was left to the team leader. Shortly thereafter the physicist returned to the farm in order to explain the report to the farmer. He asked for a blackboard and then drew a circle. He began: "Assume the cow is a sphere . . ."[1]

A common attitude among many philosophers is that science has a special epistemic status—a privileged source of knowledge—and supports realism, broadly construed, about the existence of a mind-independent world and (or) objects within it.[2] Often, this takes the form of scientific realism with commitments varying between metaphysical, semantic, epistemic and axiological

---

[1] "Old" physicist's/mathematician's joke. Exact origins are unknown to me but it is presented in a delightful version in Lawrence M. Krauss' (1994, 3-4) *Fear of Physics*.

[2] See, for example, Khlentzos (2011) (who emphasizes metaphysical realism):

> Many philosophers believe metaphysical realism is just plain common sense. Others believe it to be a direct implication of modern science, which paints humans as fallible creatures adrift in an inhospitable world not of their making.

Similar sentiments can be extracted from other authoritative survey articles such as Miller (2010) and Chakravartty (2011).

dimensions.[3] That said, even those that support non-realist views regarding science[4] share some basic tenets with the scientific realist. Arthur Fine (1996, 128) explains:

> They must both accept certified results of sciences as on par with more homely and familiarly supported claims … both realist and antirealist accept the results of scientific investigation as "true" …

Accordingly, it ought to be worrisome to many that our best scientific theories indulge in ubiquitous appeals to falsehoods, distortions and unfaithful representations of reality for the purpose of attaining empirical adequacy and explanatory success.[5] One notable line of argumentation, emphasized repeatedly by Cartwright (1983, 1989, 1999), is that the use of idealizations, abstractions, and approximations strongly undermines a realist conception of science, and can be taken as grounds for both denying the truth of scientific assertions and questioning our ability to empirically confirm scientific theories. Crudely put, the motto is that our most mature scientific theories and best accounts of physical phenomena, akin to their superseded predecessors, are "strictly speaking, false." But perhaps a more apt characterization of the state of affairs may be stated as follows: Our best scientific accounts are highly idealized,

---

[3] Roughly, the metaphysical dimension has to do with commitment to the existence of a mind-independent world and/or objects within it, (possibly) along with their properties and relations. The semantic dimension concerns interpreting the claims of scientific theories literally, as ones satisfying truth conditions. The epistemic dimension regards the empirical and explanatory success of science as evidence for the (approximate) truth of the claims of science, so that scientific claims constitute knowledge about the world and objects within it. The axiological dimension takes the aim of science to give approximately true descriptions and faithful representations of the world. See Boyd (1983, 45), Psillos (1999, xix), Niiniluoto (1999, 21) and Chakravartty (2011). See Balaguer (1998) and Leng (2010, 2012) for the type of nominalistic scientific realism that will be assessed in Chapter 9.

[4] E.g., constructive empiricism (van Fraassen 1980), historicism (Kuhn 1970, 1983), pragmatism (Peirce 1998).

[5] My tendency is to use "scientific accounts" as an umbrella term for scientific theories, laws, models, explanations, descriptions, representations, etc., wherein the process of "accounting" might take the form of an explanation (on one's favorite sense of the notion), derivation, deduction, prediction and retrodiciton, description, representation, or just a general sense of insightfulness and illumination.

abstracted, and approximated misrepresentations of reality.[6] What then justifies the commonsensical notion that science has a special access to knowledge and that it is a rational and coherent enterprise, let alone that it supports some version or other of realism?

A common attitude shared by philosophers and scientists alike, and meant to alleviate the woes associated with appeals to idealizations, is captured by the following statement made by John Earman (2004, 191):

> EARMAN'S SOUND PRINCIPLE[7] — While idealizations are useful and, perhaps, even essential to progress in physics, a sound principle of interpretation would seem to be that no effect can be counted as a genuine physical effect if it disappears when the idealizations are removed.[8]

In other words, we can view idealizations as scaffoldings of sorts. They are temporary structures there for mathematical convenience, practical or instrumental purposes, and possibly for reasons of pedagogy. That being said, in principle, they can be removed. They are not part and parcel of mature scientific theories. They do not represent genuine physical effects, and hence they cannot be the basis of a scientific account of some physical phenomenon.

---

[6] Unless otherwise qualified, I shall use the term "idealization" to refer to idealizations, abstractions, approximations, and misrepresentations. See Chapter 4 for a literature review concerning these notions, as well as an identification of how they are characterized and distinguished in the literature.

[7] To my best knowledge, Jones (2006, 194) is the first to call (a version of) this "Earman's Principle" and Ruetsche (2011, 336) the first to call it the "Sound Principle."

[8] Similar sentiments, among many others, arise in Laymon (1985, 1989), McMullin (1985, 257), Nowak (1980), and in Galileo (Drake 1967, 117 and 225). See Chapter 4 for details. Earman elaborates (in conversation): More cautiously, if a theory uses an idealization to predict an effect which disappears when the idealization is removed then either the effect is an artifact of the idealization or else (if experiment confirms the effect) the theory is inadequate. Ruetsche (2011, 336) adds: "No effect predicted by a non-final theory can be counted as a genuine physical effect if it disappears) and stays disappeared from that theory's successors."

However, one wonders whether there are idealized scientific representations, or misrepresentations, that do not fit neatly into the story told above. Representations that, when de-idealized, render the scientific account that appeals to them ultimately unsuccessful. Questions arise: Do such "essential idealizations" and misrepresentations exist? Do canonical accounts of scientific representation and idealizations address such phenomena? What would be a justification for appealing to such idealizations? For instance, can one characterize essential idealization in a manner that conforms to Earman's sound principle (or, "the sound principle," for short)? Should essentially idealized theories, laws, or models be considered explanatory or descriptive? Can idealizations be necessary in accounting for physical phenomena? If idealizations are, in some sense, dispensable, are there any substantive roles for idealizations to play in science? Are idealizations there simply for instrumental reasons, e.g., to make the math easier?

The purpose of this dissertation is to explore and suggest answers to such questions. Or, said differently, I aim to make headway on solving the following problem:

THE ESSENTIAL IDEAIZATION PROBLEM (EIP) — We need an account of how our abstract and essentially idealized scientific representations correspond to the concrete systems observed in the world, we need a characterization of essential idealizations, and a justification for appealing to such idealizations, i.e., an explanation of why and which indispensable idealizations are successful.

In a nutshell, my proposed solution will be to endorse what I will call a content-based account of representation, to justify certain essential idealizations instrumentally, reject others as pathological, but still identify substantive explanatory, methodological, and pedagogical roles for idealizations to play in science, thereby defending a compatibilist approach to the role of idealizations in science—namely, a midway position between those who propose to dispense with idealizations and those who believe idealizations are essential to scientific accounts of physical phenomena.

In particular, I'll begin in Chapter 2 by looking to the concrete case study of phase transitions to identify and flesh out the EIP. Chapter 3 will concentrate on the *representation* component of the EIP, namely, the representational relation between theory and world. I will conduct an interdisciplinary comparative analysis and argue for a content-based approach to scientific representation. The connection between Chapter 3 and the proposed solution to the EIP is as follows: By endorsing a content-based account of representation, one is committed to the idea that theoretical investigations are needed in order to decipher the contents of our scientific representations (theories, models), but such investigations include appealing to idealizations and abstractions that allow for an exploration of the possible structure and representational capacities of a theory. Chapters 4 and 5 will treat the *idealization* component of the EIP. In Chapter 4 I will look to the canonical literature on idealizations and argue that, indeed, the sound principle is both the standard and most plausible justification for appealing to idealizations, and Chapter 5 will build on prior work in order to introduce the notion of a *pathological idealization*, i.e., an idealization that is absolutely inconsistent with the sound principle. Subsequently, Chapters 6-7 put distinctions and concepts introduced in prior chapters into action, so to speak, in some concrete case studies including the Aharonov-Bohm effect and the fractional quantum statistics allegedly manifested in the quantum Hall effects. In Chapter 8, I revisit the question of whether or not idealizations can be essential to scientific accounts of physical phenomena and defend my compatibilist approach via an identification of the substantive roles that idealizations play in science (including pedagogical, methodological, and explanatory roles) and a characterization of essential idealizations based on core tenets endorsed by both camps of the essential idealization debate.

At this point the EIP will be solved, or, at least, substantial headway will have been made in solving the EIP and pathways for future study will have been identified. In Chapter 9 I attempt to reconnect with the issue of realism by outlining some implications of my work for the realism/anti-realism debate in the philosophy of mathematics. I end the dissertation with a short conclusion in Chapter 10.

In essence, this dissertation is a collection of thematically connected, but independent, papers, each of which is meant to portray an important contribution to the scholarly literature. Although independent, it is fair to say that, *roughly*, all the chapters collectively suggest the following generic thesis of the dissertation:

> **Generic Thesis**: *Canonical accounts of scientific representation and idealization, fail. They do so for two main reasons. First, they are too narrow. Second, they do not do justice to important examples.*

Or, perhaps more humbly: Many people have done good work on scientific representation and idealizations, but there is lots more important work to do and this dissertation does just that. In particular, the "important examples" include those discussed above (and in

Chapters 2, 6-7, and 9), while "narrowness" is essentially used as an umbrella term for the various mini-theses argued for throughout the entire work (in Chapters 3, 4-5, and 8) and identified in the subsection that follows. More specifically, the chapters in this dissertation will work collectively to argue for the following thesis:

> **Specific Thesis**: *The sound principle is the standard and most plausible justification for appealing to idealization in science, and this implies that pathological idealization ought to be dispensed with in accounting for concrete phenomena in the world. However, and first, pathological idealizations still play indispensable pedagogical and methodological roles in science. Second, non-pathological idealizations remain essential for particular types of explanatory purposes.*

Further details will be given in the remaining subsection of this chapter, where I present a chapter-by-chapter summary of the main content and claims of the dissertation.

1.1 CHAPTER-BY-CHAPTER SUMMARY

Chapter 2 begins by discussing the so-called paradox of phase transitions. I look at the philosophical literature on phase transitions, with an attempt to understand what the "paradox" is. I identify how almost all philosophical accounts of phase transitions and related claims, e.g., that phase transitions are examples of emergent phenomena, stem from the paradox, and so I set a condition of adequacy on any such account of phase transitions: that there really must be a bona fide paradox of, i.e., a contradiction corresponding to, phase transitions (Section 2.1). I then continue to argue for a negative and positive thesis via the distinction between abstract and concrete objects.

My negative thesis is that there really is no paradox of phase transitions and that in order to get a bona fide paradox one must undertake substantial philosophical work and ground a type of *indispensability argument*, akin to the kind appearing within the context of the philosophy of mathematics (Section 2.2). Since none of the proponents of the phase transitions debate undertake such work, and since indispensability arguments are highly controversial, I claim that the entirety of the debate, insofar as it is grounded in the paradox of phase transitions, is misguided and that the philosophical import that has been extracted from the case study of phase transitions with regard to emergence, reduction, explanation, etc., is not warranted.

My positive thesis is to show how the "paradox" can be generalized and arises whenever a scientific account appeals to an essential idealization coupled with the notion of a faithful representation (Section 2.3). Thus, I suggest that what is really interesting about phase transitions is the manner by which they illustrate the EIP, which is tightly connected to issues arising in the context of scientific representation and scientific realism. The upshot is that, insofar as proponents of the phase transitions debate have been contributing to this problem, certain aspects of the debate have been fruitful. The rest of the dissertation is dedicated to solving the EIP by concentrating on two of its main components, representation and idealization, and fleshing out some of the philosophical consequences of my work.

Chapter 3 tackles the notion of scientific representation. There are two worthwhile points to note. (i) Most discussion of scientific representation *per se* focuses on the *constitution question*, which asks for the necessary and sufficient conditions for a vehicle of representation V to represent a target of representation T. (ii) There is a (rough) consensus that the main tension in the literature arises between what one may call *functional theories*, which emphasize uses in cognitive activities (such as inferential practices), and *informational theories*, which emphasize objective relations (such as an isomorphism) between vehicle and target.

In contrast, (i*) my analysis shows that the deep problem of scientific representation concerns answering the following two questions: (Q1) What are the *contents* of this or that representational vehicle (*content identification*) and (Q2) in virtue of what facts are such contents determined (*content determination*)? Hence, my work attempts to refocus and reorient the debate on scientific representation, calling for a substantial research program to be undertaken with respect to said questions. Furthermore, (ii*) I defend a thesis by Anjan Chakravartty (2010), which states that there is no tension between functional and informational theories, by identifying how functional theories mainly target Q1, while informational theories concentrate on Q2. My analysis leads naturally to a content-based account of scientific representation and a simple answer to the constitution question:

[cont] V represents T if and only if V's (representational/semantic) contents are about T.

My method includes showing how certain requirements must be set on any tenable account of scientific representation, such as the requirement allowing for misrepresentation (Section 3.2). I then continue to argue that two leading accounts of scientific representation—the *inferential account* (Section 3.4) and the *interpretational account* (Sections 3.5, 3.6)—are flawed for they do not satisfy such requirements. Through such criticism, and drawing on an analogy from non-scientific representation (Section 3.3), I also sketch the outline of a superior *content-based* account (Section 3.7). In particular, I propose to take *epistemic representations* to be *intentional objects* that come with *reference*, semantic *contents* and a representational *code,* and I identify *faithful representations* as representations that act as *guides to ontology.*

To emphasize, in addition to making available the terminology and concepts needed for a tenable account of scientific representation meant to accommodate the practice of idealizations in science, the work of Chapter 3 connects directly to the solution of the EIP in the following manner: By endorsing a content-based account of representation, one is committed to the idea that theoretical investigations are needed in order to decipher the contents of our scientific representations (theories, models), but such investigations include appealing to idealizations and abstractions that allow for an exploration of the possible structure and representational capacities of a theory.

Chapter 4 has two main goals. The first is to present a critical survey of the current state of the literature on idealizations. I discuss distinctions between idealizations, abstractions, and approximations (Section 4.2) and present some of the main taxonomies due to McMullin (1985), Weisberg (2007a, 2013), the Poznań School (Nowak 1980), and Shaffer (2012) (Section 4.3). The main upshot will be that indeed the sound principle is the standard and most plausible justification for appealing to idealizations in science.

The second goal is to sketch how the pervasiveness of idealizations, abstractions and approximations in science raises various foundational problems (Section 4.4). Roughly, many such problems can be defused insofar as the idealizations in question are consistent with the sound principle, namely, by appealing to some standard *de-idealization* or *concretization* scheme, in which more realistic scientific representations and models are accompanied by improvements in predictive output, and/or improvability of other theoretical merits such as explanatory and descriptive power. The punch line of this section is that insofar as de-idealization schemes fail for cases of essential idealizations, the problems identified reemerge with a vengeance, so to say.

Chapter 5 presents a distinction due to Butterfield (2011) and Norton (2012) that plays an important role in identifying the possible problems (or lack of) corresponding to essential idealizations (Section 5.1). I define and explain the significance of the notion of a *pathological idealization*, i.e., an idealization that is genuinely incompatible with the sound principle, and identify three types of possible incompatibilities (Section 5.2). Subsequently, I make use of such distinctions and concepts in the following chapters.

Chapter 6 turns to investigate the representational structures and idealizations that arise in standard accounts of the Aharonov-Bohm (AB) effect (in which an interference pattern, manifested by a beam of electrons in standard double-slit experiments gains a shift in the pattern due to presence of an isolated magnetic field). It is suggested that interpreting the effect as fundamentally topological in nature commits one to an untenable view of the necessity of idealizations in science. Specifically, the *received view* of the AB effect states that one must appeal to a topologically non-simply connected (configuration) space in order to account for the effect (Section 6.2). However, I argue that such a topological idealization is a pathological one, and hence does not conform to the standard justification via the sound principle (Section 6.3). Moreover, I submit that this undesirable consequence can be evaded by embracing an alternative non-topological approach to the nature of idealizations arising in the AB effect (Section 6.4).

Chapter 7 extends the discussion in Chapter 6 to the context of the fractional quantum statistics allegedly manifested by anyons in fractional quantum Hall effect (FQHE) systems. After a short introduction to anyons and the quantum Hall effects (Sections 7.1 and 7.2), I continue to outline two approaches to fractional quantum statistics. The standard account, *the* *topological approach*, is precise and well understood (Section 7.3). However, I argue that it appeals to pathological idealizations (as discussed in Chapters 5 and 6), since the approach takes fractional statistics to be grounded in the one-dimensional unitary representation of the fundamental group (the first homotopy group) of the configuration space of identical particles in *exactly* two-dimensions. But FQHEs are not two-dimensional. Accordingly, I suggest an alternative approach, *the geometric approach*, based on calculations done by Arovas, Schrieffer, and Wilczek (1984) to show that excited FQHE states obey fractional statistics (Section 7.4). The geometric phase approach does not stem from a solid foundation like the configuration space approach, but it also does not appeal to pathological idealizations. I then make some headway in identifying what kind of work must be done in order to develop the conceptual foundations of the geometric phase approach (Section 7.5).

Chapter 8 looks to the question at the center of the growing literature on essential idealizations, specifically, whether idealizations are genuinely necessary for scientific accounts of physical phenomena (Section 8.1). A debate has risen between those who embrace essential idealizations, call these the "essentialists," and those who abhor them, call these the "dispensabilists." The purpose of the chapter is to show that the division between essentialists and dispensabilists is in fact a false dichotomy. I do so in three steps. First, I show that even pathological idealizations have indispensable pedagogical and methodological roles to play in science (Section 8.2). In addition, I argue that the well-received account of idealizations presented in Chapter 4 does not account for, but distorts the story behind the idealizations that arise in the case studies I look at (in Chapters 6 and 7). Second, I note that certain non-pathological idealizations are essential for giving "structural," "asymptotic," or "minimal model" explanations of physical phenomena (Section 8.3). Hence, my slogan:

> *Idealizations are essential to explanation and representation, as well as to methodology and pedagogy, but they essentially misrepresent.*

Third, I propose a working characterization of essential idealizations based on insights offered by both camps and in doing so contend that essentialist and dispensabilists views are importantly complementary (Section 8.4). I also sketch some examples intended to implement and further elucidate the characterization (Section 8.5). In short, my working characterization runs as follows:

> Essential Idealization — Novel and robust mathematical structure that arises via a Nortonian approximation (or idealization), secludes those features that are relevant for soundly representing and/or asymptotically explaining phenomena of interest, and is essential for the success of present science and/or will underlie the empirical success of future theories.

Together, the three steps constitute my own attempt to solve, or make substantial headway into solving, the EIP posed above and in Chapter 2 (Section 8.6). Possible objections from both the essentialist camp (Section 8.3.1) and the dispensabilist camp (Section 8.6.1) are considered.

Chapter 9 attempts to place my discussion of essential and pathological idealizations in its broader context by entering into a recent debate regarding whether or not there exists a so-called "easy road" to nominalism. Specifically, the question motivating (what I will call) the *easy* *road nominalism* debate is whether or not it is possible to reject platonism about mathematical objects, while committing to some substantial form of scientific realism, but doing so without taking the "hard road" to nominalism (Section 9.1). The hard road includes purging our best scientific theories from quantifying over abstract mathematical objects. One prominent and promising attempt to take the easy road includes Leng's (2010, 2012) approach, which defends a version of *nominalistic scientific realism*, and depends on the idea that physical structure can "approximately instantiate" mathematical structure (Section 9.2). Against Leng's approach, I show that standard topological approaches to fractional statistics and the AB effect block her path to easy road nominalism (Section 9.3). The reason is that on such approaches there is no sense in which the mathematical structure appealed to in order to explain fractional statistics and the AB effect is "approximately instantiated" in a physical system. Accordingly, I consider a nominalist rejoinder, which emphasizes that on my alternative non-topological accounts to these effects, Leng's approach to easy road nominalism becomes a viable option once more (Section 9.4). The main goal is to show how discussions of essential and pathological idealizations have consequences for the easy road nominalism debate, and so I suggest that proponents of the debate pay closer attention to such case studies from science (Section 9.5).

I end the dissertation with some final concluding remarks in Chapter 10.

2.0 THE PARADOX OF PHASE TRANSITIONS

As a motivation of sorts, this chapter looks to the concrete study of phase transitions and the philosophical debate that revolves around phase transitions in order to identify and explicate the Essential Idealization Problem (EIP). Subsequently, Chapters 3-8 will attempt to make substantial headway in solving this problem.

2.1 WHAT IS THE "PARADOX OF PHASE TRANSITIONS"?

"Phase Transitions" (PT) include a wide variety of common and not so common phenomena in which the qualitative macroscopic properties of a system or a substance change abruptly. Such phenomena include, among others, water freezing into ice or boiling into air, iron magnetizing, graphite spontaneously converting into diamond, and a semi-conductor transitioning into a superconductor. There exists a flourishing scholarly debate with respect to the philosophical import one should infer from the scientific accounts of phase transitions, in particular the accounts' appeal to the "thermodynamic limit" (TDL), and regarding how the nature of PT is best understood. It has become standard practice to quote the authoritative physicist, Leo P. Kadanoff, who is responsible for much of the advances in Renormalization Group methods and in understanding PT, in order to better illustrate the puzzlement associated with PT:

> The existence of a phase transition requires an infinite system. No phase transitions occur in systems with a finite number of degrees of freedom. (Kadanoff 2000, 238)

If we add to the above that observations of boiling kettles confirm that finite systems do undergo PT, we conclude that a rather odd paradox arises: PT do and do not occur in finite, and thus concrete and physical, systems. The above is taken as a basis for warranting such scholarly claims to the effect that PT are irreducible emergent phenomena (e.g., Lebowitz 1999, S346; Liu 1999, S92; Morrison 2012, 143; Prigogine 1997, 45), which necessitate the development of a new physical theory (Callender 2001, 550), and for inducing a wide array of literature that argues to the contrary (e.g., Bangu 2009; Batterman 2005; Butterfield 2011; Menon and Callender 2013; Norton 2012; Wayne 2009).

In this section, I would like to build on the works of Mainwood (2006) and Jones (2006) to further investigate what exactly is the "paradox" of PT, which is meant to license the type of scholarly conclusions and discussions noted above. It seems to me that a natural condition of adequacy for the particular claim that PT are emergent phenomena, as well as the more general debate that arises, is that there really is a bona fide paradox associated with PT. In other words, it really must be the case that a phase transition "is emergent precisely because it is a property of finite systems and yet only reducible to micro-properties of infinite systems" (Liu 1999, p. S104), or more recently, that "the phenomenon of a phase transition, as described by classic thermodynamics cannot be derived unless one assumes that the system under study is infinite" (Bangu 2009, 488).[9] Accordingly, in Section 2.2, I describe the paradox and suggest that much

[9] Even more recently: "A well-known fact about phase transitions is that even though they take place in finite systems, they can be accounted for only by invoking the thermodynamic limit $N \rightarrow \infty$… this happens only in infinite systems…" (Morrison 2012, 156-158).

of the debate revolving around PT stems from it. In doing so, I appeal to Contessa's (2007, 52-55) distinction between "representation" understood as "denotation," and "faithful representation" understood as a type of "guide to ontology" (Sklar 2003, 427).[10] Afterwards, I will continue to argue for a negative and a positive thesis. My negative thesis is that there really is no paradox of phase transitions and that in order to get a bona fide paradox, i.e., a contradiction, one must undertake substantial philosophical work and ground a type of *indispensability argument*, akin to the kind appearing within the context of the philosophy of mathematics.[11] Since none of the proponents of the PT debate undertake such work, and since indispensability arguments are highly controversial, I claim that the entirety of the debate, insofar as it is grounded in the paradox of PT, is misguided and that the philosophical import that has been extracted from the case study of PT with regard to emergence, reduction, explanation, etc., is not warranted.

However, I also have a positive thesis. In Section 2.3 I show how the "paradox" can be generalized and arises whenever a scientific account appeals to an "essential idealization"[12]—roughly, when a scientific account of some concrete physical phenomena appeals to an idealization in which, in principle, one cannot attain a more successful account of said phenomena by "de-idealizing" the idealization and producing a more realistic idealization. In doing so, I suggest that what is really interesting about phase transitions is the manner by which they illustrate the "Essential Idealization Problem" (EIP), which is tightly connected to issues arising in the context of scientific representation and scientific realism. The upshot is that, insofar as proponents of the phase transition debate have been contributing to the EIP, certain

---

[10] See Chapter 3 for details on "faithful representation."
[11] See Colyvan (2001, 2015) for more on indispensability arguments and a defense.
[12] Butterfield (2011) and Mainwood (2006) use the term "Indispensible," Jones (2006) uses "Ineliminable," and Batterman (2005, 2013) uses "Essential."

aspects of the debate have been fruitful. My own contribution to the solution of the problem appears in the following chapters (with an emphasis on Chapter 8).

Before continuing and diving into the paradox of phase transitions, it would do well to give a schematic presentation of the theoretical framework in which phase transitions are studied. I refer the reader to Kadanoff (2000) and Stanley (1971) for standard textbooks, and to Batterman (2002, 2005) and Butterfield (2011, Section 7) for philosophical friendly accounts. Here I only sketch the framework.

Canonical accounts of phase transitions characterize them by non-analyticities (or discontinuities) in the partition function (per particle) associated with some system, or (similarly) as discontinuities in the various thermodynamic potentials associated with the system such as the (Helmholtz or Gibbs) free energy (see Figure 2.1). In this sense non-analytic partition functions represent phase transitions since they mark the occurrence of a concrete phase transition (as might appear in a lab). Moreover, such representational structure is faithful in the sense that it allows for a great deal of sound inferences to be made about an actual system undergoing a phase transition.

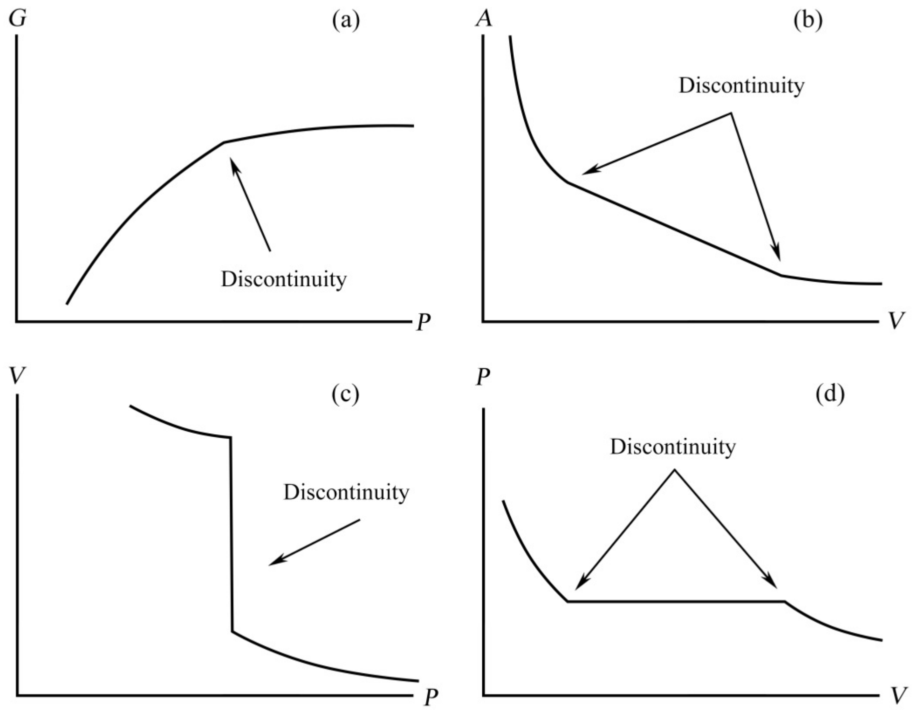


Figure 2.1: Graphs Displaying a First-order Phase Transition. Graph (a) displays the Gibbs free energy (or Gibbs thermodynamic potential) G as a function of the pressure P, graph (b) displays the Helmholtz free energy (or Helmholtz thermodynamic potential) A as a function of the volume V. Graphs (c) and (d) display functional relations between P and V. Based on Stanley (1971, 31).

For instance, concentrating on Figure 2.1, one can establish the various functional relations between the thermodynamic potentials associated with some system (which can be ascertained by taking partial derivatives of the system's partition function), and properties such as pressure, volume and temperature. However, known methods for making such mathematical structure available all appeal to the TDL in which the system's particle number and volume diverges. To get a sense for why this is the case, note that a partition function is a sum of nondiscontinuous analytic functions, but any such finite sum will also be analytic. Yet, such a result can be avoided by appealing to a thermodynamic-type limit. For instance, the series $\sum_{n=0}^{N} x^n = 1 + x + x^2 + \cdots + x^N$ has no discontinuity for any finite sum. But if we allow for infinite sum $N \to \infty$ the series tends toward $\frac{1}{1-x}$ and has a discontinuity at $x = 1$. We need to take the limit in order to get the discontinuity. Analogously, we need to take the TDL in order to allow for non-analytic partition functions to arise.

2.2 IDENTIFYING THE "PARADOX" OF PHASE TRANSITIONS

In his 2001 paper, "Taking Thermodynamics Too Seriously," Craig Callender presents several allegedly true propositions that jointly induce a paradox concerning PT—that concrete systems can and cannot undergo PT:[13]

1. Concrete systems are composed of finitely many particles N.
2. Concrete systems display PT.
3. PT occur if and only if the partition function Z has a discontinuity.
4. The partition function Z of a system with finitely many particles N can only display a discontinuity by appealing to the TDL.
5. A system in the TDL has infinitely many particles.[14]

[13] The paradox of PT presented here is not the exact version presented in Callender (2001, 549). Instead, I present the paradox in a manner that is more relevant to my discussion. Several authors, such as Mainwood (2006, 223) and Jones (2006, 114-7), have undertaken a similar approach.
[14] For precise characterization of various forms of the TDL, see Norton (2012, Sections 3 and 4) and references therein.

Tenets 1-2 imply that concrete and *finite* systems display phase transitions while tenets 3-5 imply that only *infinite* systems can undergo a phase transitions. However, *contra* Bangu (2009), Callender (2001), Mainwood (2006), Jones (2006) and others, I contend that no contradiction arises by conjoining tenets 1-5. To see this, we must first distinguish between "concrete" PT, on the one hand, and "abstract mathematical representations" of them, on the other hand.[15] To be clear, a "concrete" system would include a physical thermal system of the type we find in the world or in a lab, while "abstract mathematical" just refers to pieces of math, e.g., a set with a function defined on it.[16] Also, I take the term "representation" here to be stipulated denotation that is agreed upon by convention.[17] For instance, the notation "N" represents "the number of particles" (in a given system) in the sense that it *denotes* the number of particles. Second, notice that there are ambiguities with regard to whether the terms "PT" and "partition function" ("Z") in tenets 3 and 4 refer to concrete objects, or abstracts mathematical representations of them. As *concrete objects*, PT are concrete phenomena or processes that arise within concrete systems, while Z is some sort of concrete property of such systems. As *abstract mathematical representations*, both PT and Z are just pieces of mathematics that allegedly denote concrete objects. To avoid confusion, note that by "abstract PT" I only mean PT in the sense that an abstract Z displays a discontinuity. In the same manner, there is a clear ambiguity concerning the physical interpretation, i.e., the concreteness or abstractness, of the TDL.

---

[15] The distinction between concrete and abstract objects is well-known. Abstract objects differ from concrete ones in the sense that they are non-spatiotemporal and causally inefficacious. Paradigm examples include mathematical objects and universals. See Rosen (2001).

[16] I remain agnostic regarding the possible mind-independent existence of such entities (Platonism), versus their mind-dependent existence as fictions of sorts (nominalism). Chapter 9 will bring such issues to the forefront.

[17] See Contessa (2007, 52-55) and references therein.

Thus, for example, if "PT" and "Z" in tenets 3 and 4 refer to abstract mathematical representations, as opposed to concrete objects, then there is no paradox: Concrete and finite systems display PT while abstract and finite ones do not. Just because abstract mathematical representations of concrete systems with finite N do not display PT, this does not mean that concrete finite systems do not display PT. Alternatively, if "PT" in tenets 3 and 4 do refer to concrete PT, it also does not immediately follow that there is a paradox. Rather, what follows is that concrete PT "occur" when abstract representations of them display various abstract properties, such as a discontinuity in Z and an appeal to the TDL. One might wonder what explains this particular correlation between discontinuities in abstract representational partition functions and concrete phase transitions. However, *prima facie*, there is no paradox.

The point is that without adding additional tenets that make a claim about the relation between, on the one hand, concrete PT occurring in physical systems and, on the other hand, the abstract mathematical representation of concrete PT, which arise in scientific accounts of PT, no paradox occurs. In the following subsection I will add such additional tenets in hope to further shed light on the central philosophical issue that arises in the context of PT. To end, it is worth noting that, if my claim about there being no paradox is correct, then the entire debate revolving around PT, insofar as it is grounded in the paradox of PT as stated above, is unmotivated and misguided. In particular, notice that the various positions expressed with regards to the debate can be delineated by identifying which tenet of the paradox is denied or embraced by a particular proponent. Authors such as Lebowitz (1999, S346), Liu (1999b, S92), Morrison (2012, 143) and Prigogine (1997, 45) can be read as embracing tenet 3 and identifying PT as a kind of non-reductive emergent phenomenon. Contrasting attitudes have been voiced by Wayne (2009), where Callender (2001) and Menon and Callender (2013) explicitly deny that phase transitions are irreducible and emergent phenomena by rejecting tenet 3. Butterfield (2011) can be read as both denying and embracing tenet 3, in an effort to reconcile reduction and emergence. Norton (2012) can be understood as denying tenet 5.[18] I refer the reader to Mainwood (2006, 223-237), who presents an exposition of this type of delineation—i.e. a classification of scholarly attitudes to the nature of phase transition grounded in the paradox. For my purposes what is important is to identify that *the large majority, if not all, of the phase transition debate arises from the alleged paradox of phase transition.*

2.2.1 The Bona Fide Paradox of Phase Transitions and its Generalization

The key ingredient necessary to engender a bona fide paradox is for a particular kind of correspondence relation to hold between abstract representations and concrete systems. To make this point clear we must appeal to a further distinction. While I take "representation" to be stipulated denotation, by "faithful representation" I mean a representation that allows agents to perform sound inferences from the representational vehicle to the target of representation (Contessa 2007, 52-55). That is to say, a faithful representation allows agents to make inferences about the nature of the target of representation. Thus, it acts as a kind of "guide to ontology" since it accurately describes aspects of the target of representation (Sklar 2003, 425). In other words, a faithful representation is one in which the vehicle and target of representation resemble each other in some manner, e.g., they share some of the same, or approximately same, properties and/or relations. The classic example here is a city-map, which is a faithful representation of a

[18] See Norton (2012) for details. In short, Norton differentiates between a strong version of the TDL, in which reference to an infinite system is presumed, and a weak version of the TDL, in which the limiting procedure corresponds to the behavior of certain sequences of properties and thus there is no reference to an infinite system. Subsequently, he urges us to dispense with the strong version and appeal only to the weak version in accounting for phase transitions.

city because it allows us to perform sounds inferences from the vehicle to the target, i.e., from the map to the city. This is so because both the vehicle and the target share various properties. For instance, if two streets intersect in the map, then they also intersect in the city. That is to say, intersecting streets in the map correspond to intersecting streets in the city. Therefore, the map acts as a type of ontological guide accurately describing the city, e.g., there *really are* intersecting streets in the city. It is worth noting that my account here differs from Contessa (2007), who isn't clear about the ontological aspect of faithful representations.[19] Contessa (2007) differentiates between "epistemic representations," from which *valid* inferences can be drawn, and faithful ones that permit *sound* inferences. Whether or not such inferences come with ontological baggage depends on whether they are about the target itself. On my account here, faithful representations license sound inferences about the target itself and hence they the fix the ontology of the target.[20]

With this distinction in hand, if we add a tenet which states that the abstract representational discontinuities representing phase transitions are *faithful* and hence correspond to concrete physical discontinuities we do get a genuine contradiction. This is so because if systems are composed of finitely many particles, which is the case within the context of the atomistic theory of matter conveyed in tenet 2, then it makes no sense to talk of concrete discontinuities. The notion of concrete discontinuities presupposes that matter is a continuum so that there can be an actual discontinuity. Otherwise, an apparent discontinuity is actually the rapid coming apart of particles and not a real discontinuity. Consequently, adding a tenet as the one just described amounts to claiming that systems are not composed of finitely many particles

---

[19] I treat the issue in Section 3.5.

[20] See Chapter 3 for a more thorough and detailed discussion. It turns out that the relation between faithful representations as surrogates for sound inferences, on the one hand, and guides to ontology, on the other, is a complicated subject that deserves more attention than what can be given in this chapter.

and so we arrive at the following statement: Concrete systems are and are not composed of finitely many particles N.

In a similar manner, one can engender a kind of paradox by reifying the TDL through an appropriate correspondence relation. For instance, one could add the tenet that an appeal to the TDL, which could be interpreted as a type of continuum limit faithfully representing an *abstract* system, in fact faithfully represents a *concrete* system.[21] Thus, we deduce the claim that concrete systems are and are not composed of finitely many particles N (in the sense that the ontology of concrete systems is both atomistic and that of a continuum, i.e., not atomistic).

The source of the problem of PT seems to be that the mathematical structure that scientifically represents concrete PT—a discontinuity in the partition function—looks like an artifact of an idealization (or an approximation)—the TDL—which is essential in the sense that when one "de-idealizes" said idealization, the mathematical structure representing PT no longer exists.[22] Accordingly, I would like to suggest that what is really interesting about PT is the manner by which they might shed light on the nature of scientific representation and idealization. In particular, notice that when concerns regarding representations are incorporated, the paradox of PT can be generalized by making use of the concept of an essential idealization:

1. Concrete systems include a concrete attribute A.[23]
2. Concrete systems display a concrete phenomenon P.
3. P is scientifically-mathematically represented by P'.

---

[21] But see Norton (2012) for an argument to the effect that the so-called "continuum limit" does not correspond to a continuum and, in the terminology introduced in Chapter 5, is a case in which no limit system exists.
[22] For a more precise statement, see Butterfield's (2011, 1123-1130) and Mainwood's (2006, 216-218) discussion of Yang-Lee Theory and KMS states.
[23] By "attribute" here, I just mean "property," or a relation of some order. I only use attribute, denoted by A, so as note to confuse it with a phenomenon, denoted by P.

4. P' can only arise by appealing to an idealizing limit I.
5. A system in the idealizing limit I includes an attribute $A^{\approx}$ such that $A \neq A^{\approx}$.
6. P' faithfully represents P.

Tenets 1 and 2 imply that concrete systems are A and display P. Tenets 3-5 imply that P is scientifically represented by P', which presupposes $A^{\approx}$. Tenet 4 encompasses our essential idealization since any de-idealization of I will render P' nonexistent. So far there is no contradiction. But, when one adds the correspondence relation described by tenet 6, a bona fide paradox arises: Concrete systems are and are not A (since they are A and they are $A^{\approx}$and $A \neq A^{\approx}$). What is important to notice is that tenets 1 and 2 are claims about *concrete* systems, wherein tenet 2 identifies the concrete phenomenon to be scientifically accounted for, while tenets 3-5 are claims about *abstract* scientific accounts of concrete systems, and it is tenet 6 that connects the abstract with the concrete via faithful representation, thereby engendering a genuine paradox.

For illustration purposes, consider how the above general paradox template can be instantiated for a case different from phase transitions (one which we study in Chapter 7):

1. Concrete systems are three-dimensional (3D).
2. Concrete systems display the fractional quantum Hall effect (FQHE) wherein concrete fractionally-charged quasi-particles are observed in an electron gas.
3. Concrete fractionally-charged quasi-particles are represented by anyons.

4. The two-dimensional (2D) limit is an essential idealization with respect to anyons in the sense that anyons only arise in 2D systems and any de-idealization of the 2D limit renders anyons nonexistent.
5. A system in the 2D limit is 2D and not 3D.
6. Anyons faithfully represent the concrete fractionally-charged quasi-particles that arise in the FQHE.

It follows that concrete systems that display the FQHE are and are not 3D (since they are 3D and they are 2D and 3D≠2D).

The question, of course, is why one would endorse tenet 6. The answer is that without tenet 6 the entire scientific account of the concrete phenomenon in question seems somewhat mysterious to anyone with non-instrumental sympathies. In particular, those with realist intuitions will want to unveil the mystery with a correspondence relation that tells us that our abstract scientific accounts gets something right about the concrete world. But how would one argue for a correspondence relation along the lines of 6? It seems to me that, given the "essentialness" aspect of the idealizing limit that arises in tenets 3 and 4, the only way to justify tenet 6 is by an appeal to an *indispensability argument*.[24] In other words, an argument to the effect that we should be ontologically committed to the existence of objects indispensable to our best scientific theories. Said differently, and in the specific cases of PT, since reference to a discontinuity in Z is indispensable to scientific accounts of PT, and since these discontinuities only arise by appealing to essential idealizations, we ought to believe in the existence of concrete discontinuities. In the following chapters I discuss the issue in detail (especially Chapters 8 and 9). For now, all that is

[24] See Colyvan (2001, 2015).

important for my account is that since indispensability arguments are non-trivial, and seem necessary to support tenet 6, which effectively gives rise to a paradox, it is fair to say that there is no paradox of phase transition that can license the type of debate undertaken in the scholarly literature.

Said differently, in contrast to many of the scholars engaged in the phase transition debate, who assume that there is a paradox and then continue to attempt to dissolve it by some manner or other, I claim that in order to get a genuine paradox one needs to justify a correspondence relation (such as the one appearing in tenet 6) by appealing to an indispensability-type argument. Since cogent indispensability-type arguments require serious philosophical work and are very much controversial, and since no author engaged in the phase transition debate has undertaken such work, it follows that much of the controversy revolving around phase transitions is not well-motivated. That is to say, claims to the effect (i) that PT are or are not emergent, (ii) that they are or are not reducible to Statistical Mechanics (SM), and (iii) that they do or do not refute the atomic theory of matter, are grounded in a frail foundation that does not license such significant conclusions.

One might worry that, contrary to my claims, a bona fide paradox of PT can arise on the epistemological level by conceding to a set of tenets from which it is possible to deduce that SM does and does not govern phase transitions. The idea here is to argue that "SM-proper" is not licensed to appeal to the TDL and so SM-proper does not govern PT. However, the objection continues, it is generally assumed that SM is the fundamental theory that governs PT. Thus, we have a paradox and the natural manner by which to dissolve it is to argue that SM-proper does indeed have the tools to account for PT (Callender 2001, Menon and Callender 2013), or else to claim that PT are emergent. In reply, it is far from clear to me that SM-proper is not licensed to appeal to the TDL, and so that it does not govern PT. In fact, there are reasons to think that the TDL is part and parcel of SM-proper because (a) it is common practice to appeal to the TDL in modern approaches to SM, and (b) the TDL is used in SM not only to account for phase transitions but to account for, among others, the equivalence of SM ensembles, the extensivity of extensive thermodynamic parameters, Bose condensation, etc. (Styer 2004). In addition, (c) all the best scientific accounts of PT, and these include mean field theories, Landau's approach, Yang-Lee theory and Renormalization Group methods, represent PT as discontinuities by appealing to the TDL, and (d) the large majority of empirically confirmed predictions of SM, within the context of PT and beyond, appeal to the TDL.

Moreover, even if it were the case that SM-proper is not licensed to appeal to the TDL, no contradiction would arise. Rather, it would be a brute fact that SM-proper does not govern phase transitions and "SM-with-the-TDL" does. If then it is claimed that the ontologies of SM-proper and SM-with-the-TDL are radically different so that indeed there is a paradox, we must notice that such a claim amounts to no more than reviving the paradox at the level of ontology, and hence my discussion in this section bears negatively on this claim.

Last, the claim that PT are emergent because SM-proper cannot account for them seems to replace one problem—PT are not governed by the fundamental theory—with another problem—PT are emergent. How does dubbing PT "emergent" illuminate our understanding of them or of their scientific accounts? How is this philosophically insightful? Accordingly, for now I endorse Butterfield's (2011) description of emergence as novel and robust behavior, as opposed to a failure of intertheoretic reduction of some sort.[25]

---

[25] In Chapter 8 I will suggest that Butterfield's (2011) notion of emergence is better saved for a characterization of EI.

2.3 THE ESSENTIAL IDEALIZATION PROBLEM

The above discussion points to what I consider to be the central philosophical issues arising out of the debate concerning PT: the discussion regarding (i) the need for a correspondence relation between our abstract scientific-mathematical representations and concrete systems, (ii) the appeal to the concept of "faithful representation," and (iii) the identification that the phase transition paradox can be generalized to any scientific account that appeals to essential idealizations, demonstrates that a solution to the following problem is needed:

> THE ESSENTIAL IDEAIZATION PROBLEM (EIP) — We need an account of how our abstract and essentially idealized scientific representations correspond to the concrete systems observed in the world, we need a characterization of essential idealizations, and a justification for appealing to such idealizations, i.e., an explanation of why and which indispensable idealizations are successful.[26]

Insofar as solutions to the EIP can be found in the literature, they pave the road for further work to be done, but it is questionable whether they are conclusive and exhaustive.[27] My

---

[26] Mainwood (2006, 214-5) also identifies a similar problem but in a context that is different from mine, and his solution (238), endorsed by Butterfield (2011), misses the central issue discussed here.

[27] For example, potential solutions to the paradox of PT can be extracted from two recent contributions to the debate: Butterfield (2011) and Norton (2012). Butterfield (2011) grants that the TDL is "epistemically indispensable" for the emergence of the novel and robust mathematical structure that is used to represent PT, but denies that any paradox emerges because the limit is not "physically real." Using the terminology expressed here, the discontinuities in Z play a representational role but not a *faithfully* representational one. The question arises: how can unfaithful representations work so well? To that end, Butterfield (2011, Section 3) appeals to the distinction, also used by Norton (2012, Section 3), between "limit quantities" or "limit properties," i.e., the limits of properties (of a sequence of systems), and a "limit system," i.e., the system at the limit. He continues to argue that the behavior of certain observable properties of concrete finite systems, e.g., magnetization of a ferromagnet, smoothly approaches the behavior of the corresponding properties of abstract infinite systems. Moreover, it is the large N behavior, not the infinite N, which is physically real.

Norton (2012) suggest that by viewing the TDL as an "approximation"—an inexact description of a target system, instead of an "idealization"—a novel system whose properties provide inexact descriptions of a target system, we can defuse any problems that might arise. Within the context of our discussion, Norton's idea is that no paradox can arise if the TDL is an approximation since approximations do not refer to novel systems whose ontology might be drastically different from the target system's ontology, thereby engendering a paradox once we add an appropriate correspondence relation. In a similar manner to Butterfield (2011), his justification for appealing to such an approximation is pragmatic: the behavior of the non-analytic Z belonging to an infinite system, is approached by an analytic Z corresponding to a finite system with large N.

From my viewpoint, although both Butterfield (2011) and Norton (2012) make substantial headway in solving the EIP vis-à-vis phase transitions, this cannot be the whole story. First, both accounts seem to ignore that it is a mathematical structure that arises only in the limit that is doing the representational work for us. Moreover, the accounts seem to suggest that we must revise our definition of PT as occurring if and only if the partition function has a discontinuity and substitute it with something along the lines of "PT occurs when various thermodynamic potentials portray sufficiently extreme gradients." The weakness of this suggestion is that we have substituted a precise characterization of PT with a vague one. But more problematic is the idea that we should be able to construct a finite N system that has, say, a Helmholtz free energy with an extreme gradient, which does not evolve into a discontinuity once the TDL is taken. Second, the Butterfield-Norton approach outlined above seems incomplete, *for it does not leave room for a substantive positive role for idealization to play in science*. Instead, Butterfield (2011) and Norton (2012) seem to think that idealizations simply ought to be dispensed with. As I will argue in Chapter 8, I don't think this is the case.

Another possible solution to the EIP, which I reject, can be extracted from Jones (2006) and from Liu (1999a, 2001 and 2004a). Jones' (2006) solution to the paradox is to deny that the TDL, or the "$N \to \infty$" syntax, should be physically interpreted as a system in the continuum or infinite particle limit. That is to say, Jones denies that the TDL is an idealization. Rather, he suggests that it be interpreted as an abstraction, in particular, as the limit in which surface and boundary effects are ignored. Since ignoring boundary effects does not amount to an outright distortion of the target system, there is no clash between the ontologies of target and analogue systems, and so, Jones claims, no paradox. Consequently, Jones urges that all cases of idealizations be re-interpreted as abstractions if possible, because abstractions appear less problematic philosophically than idealizations. One might wonder though, what justification is given for interpreting the TDL in such an unorthodox manner? The answer can be illustrated with the following example (Hill 1963, 1, 41-44): Consider the Gibbs free energy per particle $g *= g(p,T)$ for a single colloidal particle where p is pressure, T temperature, N is the number of particles of the system, G is the Gibbs free energy, $a(p,T)$ the surface free energy, and $b(T)$ and $c(p,T)$ are terms describing effects due to, among other things, rotational and translational motion (Jones 2006, 156):

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

When one takes the TDL of $g *$, the $a(p,T), b(T)$ and $c(p,T)$ terms vanish since $N \to \infty$, and g* reduces to:

$$g *= g(p,T)$$

Thus, one sees, how when taking the TDL the terms describing surface and boundary effect vanish. However, the question remains: does physically interpreting the TDL as an abstraction solve the paradox of phase transitions and can it be used to defuse the EIP? I answer in the negative for it remains a brute fact, given Jones' (2006) solution and his take on the paradox, that phase transitions occur *only* in systems where surface and boundary effects are ignored. Once one "de-abstracts" (or "concretizes") and brings such effects back in, the account of phase transitions fails. This, I submit, is just as mysterious as our original problem and so the EIP stands unsolved (only now it might be renamed the "essential abstraction problem").

Next, another manner by which to attempt to solve the paradox of phase transitions and the EIP is to argue that the TDL is not an idealization, but an approximation, and so it is justified within the mathematical formalism itself. Such a technique is considered by Liu (1999a, 2001 and 2004a), and encapsulated by the following statement by Gelfert (2005, 4, his emphasis):

> 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 of $N \sim 10^{23}$, and the statistical results

goal in the rest of the dissertation will be to make substantial headway in solving the EIP. There are two components that must be catered to in an exhaustive solution. The first concerns the representational relation between theoretical structure and the world, while the second has to do with the nature of the idealizations that give rise to the EIP. Chapters 4-8 will deal specifically with idealizations and essential idealizations, and Chapter 3, which I turn to next, will concentrate on the representational relation between theory and world.

> 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.

However, there is a worry that arises with such an approach and is explained nicely by Liu (1999, S101) so I shall quote him in length:

> [Consider small bit of mass ΔM and of volume ΔV, and let the mass density be $\sigma_M = \Delta M/\Delta V$.] … imagine a series of ΔV's , in which each later ΔV is smaller than the earlier one. At each ΔV_n, the average mass density, $\Delta M_n/\Delta V_n$, is well-defined and as $\Delta V_n \to 0$ (i.e., $n \to \infty$): $\forall \varepsilon > 0, \exists N(\varepsilon) > 0$, such that $\forall n > N(\varepsilon), |(\Delta M_n/\Delta V_n) - \sigma_M| < \varepsilon$. And since the molecules in the solid are so densely populated, the different between the actual average mass density and the limit density is so small as to make the latter a good approximation for the former… [And so we say that $\sigma_M = \lim_{\Delta V \to \infty} \Delta M/\Delta V = dM/dV$.] However, this does not seem to apply to phase transitions. The singularities [in various thermodynamic variables] are not reached asymptotically toward [the TDL], for (i) at no stage of the process in which $V, N \to \infty$, is a singularity of a system even roughly defined, (ii) nor is the singularity approached or approximated in any proper sense of approach or approximation.

In other words, it is not clear that the notion of approximation can be applied in the case of phase transitions, since the novel mathematical structure that is used to represent phase transitions—discontinuities and KMS states—arise solely in the limit. So it does not matter that, say, the Helmholtz Free Energies of a finite system and a system in the TDL look similar—one has no discontinuity and the other does, and if we take seriously the fact that phase transitions are represented by discontinuities, then approximation does not seem like the right interpretation of the TDL.

3.0 SCIENTIFIC MISREPRESENTATION AND GUIDES TO ONTOLOGY

In this chapter I argue that two leading accounts of scientific representation—the inferential account and the interpretational account—are flawed, and I also sketch the outline of a superior content-based account of scientific representation drawn from a comparative interdisciplinary analysis of the notion of representation. The content-based account holds that idealizations are indispensable for the identification and determination of the representational contents of scientific theories.

3.1 INTRODUCTION

The subject of scientific representation can be traced historically at least as far back as Charles Sanders Peirce's work on signs in the 1860's, and has received wide attention from philosophers of science in recent years (see below). Broadly speaking, I will follow Chakravartty (2010, 198-199) in identifying two main approaches to scientific representation that he calls *informational theories* and *functional theories*:

> The idea [with informational theories] is that a scientific representation is something that bears an objective relation to the thing it represents, on the basis of which it contains information regarding that aspect of the world… the other broad approach to scientific

> representation [(functional theories)] comprises theories that emphasize the *functions* of representations: their uses in cognitive activities performed by human agents in connection with their targets. The idea here is that a scientific representation is something that facilitates these sorts of activities…

Informational theories take the essence of representation to be one of a *similarity* or *resemblance* relation (Cartwright (1983), Giere (1985, 1988, Chapter 3; 1999a, 1999b, 2004), Godfrey-Smith (2006), Teller (2001), Weisberg (2013)), a *structure-preserving mapping* (of some kind) (da Costa and French (1990), French and Ladyman (1999), Swoyer (1991), Pincock (2012)), (or more specifically) an *isomorphism* (van Fraassen (1980, Chapter 3; 1989, Chapter 9), French (2003)), *homomorphism* (Bartels (2006), Lloyd (1994), Miller & Page (2007), Mundy (1986)), or *partial isomorphism* (Bueno (1997), Bueno *et al*. (2002), da Costa and French (2003, Chapter 3), French (2014)).[28] On the one hand, functional theories of representation emphasize *denotation*

---

[28] Various objections have been put forth against informational theories, e.g., Frigg (2002, 2006), Goodman (1972, 1976), Suárez (1999, 2003). For concreteness, take representation to be a similarity or resemblance relation. Similarity does not seem to be a sufficient condition on representation, since any object is similar to any other object to some extent. As Goodman (1976, 4) notes:

> None of the automobiles of an assembly line is a picture of any of the rest; and a man is not normally a representation of another man, even his twin brother. Plainly, resemblance in any degree is no sufficient condition for representation.

It has also been argued that similarity is not necessary for representation. For example—so the claim goes—words represent concepts and objects and yet there is no sense in which they are similar to their targets. More generally, anything can represent anything else through conventional stipulated denotation, and so similarity is not necessary for representation.

Another classical line of objection, originally posed by Goodman (1976, 3-10), is that the logical properties of the representation relation and the similarity relation do not match up. Similarity is a reflexive, symmetric and, arguably, transitive relation. Representation, on the other hand, is non-reflexive, non-symmetric, and non-transitive.

In addition, it has been objected that "similarity" and "resemblance" are too ambiguous to count as insightful accounts of representations. For instance, Quine (1960, 59) thought the notion of similarity is "logically repugnant," because it cannot be reduced to logical or empirical notions. We may then move on to theories that talk about representation as an isomorphism, homomorphism, or partial isomorphism between a representational structure and the world. In such a context, it is objected that either the physical world has no "structure" in the

(Goodman 1976, Hughes 1997), *exemplification* (Elgin 2004), *inferences* (Suárez 2004), *interpretation* (Contessa 2007), and *informativeness* (Bolinska 2013).

However, the general notion of representation as a subject of philosophical inquiry has a longer history and is much wider in scope, occupying a central stage in areas such as philosophy of mind and cognitive sciences, philosophy of language, and the philosophy of art. Since scientific representation is, first and foremost, an instance of representation, it seems prudent to pay attention to how the notion arises in various contexts. It is this intuition that will motivate the content-based account of representation outlined in this chapter.

Specifically, drawing again on the case of phase transitions (discussed in Chapter 2), I show how certain requirements must be set on any tenable account of scientific representation, such as the requirement allowing for misrepresentation (Section 3.2). I then continue to argue that two leading accounts of scientific representation—the inferential account (Section 3.4) and the interpretational account (Sections 3.5, 3.6)—are flawed for they do not satisfy such requirements. Through such criticism, and drawing on an analogy from non-scientific representation (Section 3.3), I also sketch the outline of a superior content-based account of scientific representation (Section 3.7). What determines the contents of scientific representations will turn out to be a substantial philosophical question meriting further study (analogous to questions about content determination in the context of linguistic and mental representations).[29]

My account amounts to a rejection of two claims commonly found in the literature: (i) The pertinent question that must be answered in a philosophical account of scientific

---

precise sense that there can be an isomorphism, etc., between representation and world, or that "structure" is also overly ambiguous.

Last, many argue that informational theories do not properly accommodate misrepresentation and idealization (Batterman 2010). I discuss this last issue further in this chapter.

[29] See Stich and Warfield (1994) for an anthology of theories of content determination in the context of mental representations. Speaks (2014) contains a thorough discussion regarding linguistic content determination.

representation *per se* concerns *the constitution question*, which asks for the necessary and sufficient conditions for a vehicle of representation V to represent a target of representation T.[30] (ii) In answering the constitution question, the main tension in the literature arises between functional theories and informational theories.[31]

In contrast, (i*) my analysis shows that the deep problem of scientific representation concerns answering the following two questions: (Q1) What are the *contents* of this or that representational vehicle (*content identification*) and (Q2) in virtue of what facts are such contents determined (*content determination*)? Hence, my work attempts to refocus and reorient the debate on scientific representation, calling for a substantial research program to be undertaken with respect to said questions. Furthermore, (ii*) I defend a thesis by Anjan Chakravartty (2010), which states that there is no tension between functional and informational theories, and that two approaches are complimentary, by identifying how functional theories mainly target Q1, while informational theories concentrate on Q2. My analysis leads naturally to a content-based account of scientific representation encompassed in a simple answer to the constitution problem:

[cont.] V represents T if and only if V's (representational/semantic) contents are about T.

3.2 SCIENTIFIC MISREPRESENTATION

In this section I would like to ascertain what requirements must be met by any acceptable account of scientific representation, such as the requirement allowing for misrepresentation. To

---

[30] See Callender and Cohen (2006), Contessa (2011), Frigg (2006), Suárez (2010).
[31] See Chakravartty (2010), Contessa (2007, 2011), Suárez (2003, 2004, 2010).

that effect, notice that scientific representation concerns scientific *methodology*. That is to say, the manner by which scientists choose to represent the world and phenomena within it, say, through scientific theories, laws, equations of motion, graphs, simulations, etc., has to do with the methods deemed best fit to satisfy the goals of science (whatever those goals might be). But there are also clear *epistemological* and *ontological* aspects to scientific representation. Scientists routinely use representational structures such as models to make inferences about the world (epistemological aspect) and to tell us what it is like (ontological aspect). An example ought to add clarity to the matter. Consider the much-debated case study of phase transitions discussed in the previous chapter. Recall, in such a context we represent phase transitions as *discontinuities* (or, synonymously for our purposes, *non-analyticities)* — i.e., points in which a function is not infinitely differentiable — in the partition function (per particle) associated with a given system, where a partition function is a function that contains information about the various microstates that a system might occupy along with their probabilities. We say that such discontinuities mark (or refer to) concrete phase transitions arising in our system of study. The choice to represent phase transition in such a manner is a methodological one. In principle, other options are also possible.[32] Nevertheless, representing phase transitions in such a manner allows us to gain much information regarding their behavior and nature. Consequently, the epistemological and ontological aspects of scientific representation enter. For instance, a quick a look at the scientific representation of phase transitions through graphical representations of a system's thermodynamic potentials (which can be ascertained by taking partial derivatives of the system's partition function) allows one to infer the functional relation between various potentials such as the relation between the Gibbs free energy and the pressure in graph (a) (see Figure 3.1). One

[32] For alternative representations of phase transitions see, for example, Gross and Votyakov (2000), Chomaz *et al*. (2001), and Borrmann *et al*. (2000).

can establish, for instance, how the pressure of a particular fluid varies as the volume is increased or decreased, and vice versa (graphs (c) and (d)). However, not every aspect of the diagrams is meant to represent. The texture of the paper on which the diagrams are printed, for example, does not represent the corresponding "texture" or "feel" of the fluid being represented. We call such non-representational properties the *artifacts* of the representation (Swoyer 1991, 463).

Moreover, even those properties of our diagrams that are meant to play a representational role, and allow us to make sound inferences about the system represented, do not necessarily tell what the system is like. Or, to use terminology introduced by Sklar (2003), certain properties of our representations are not *ontological guides*.[33] To see this, notice that all four graphs contain discontinuous jumps or kinks (see Figure 3.1). As just mentioned, these discontinuities represent the occurrence of a phase transition. However, there is an ontological sense in which the discontinuities representing phase transitions fail to represent. Or, rather, they *misrepresent* the system. It is well known that although the most sophisticated accounts of phase transitions characterize them as discontinuities (of the kind appearing in the graphs), strictly speaking, such sharp discontinuities can only arise in infinite systems (i.e., by taking some limit in which the number of particles diverges as discussed in Chapter 2). Still, neither boiling kettles nor iron bars undergoing a transition from a ferromagnetic to paramagnetic phase are infinite in extent.

[33] To clarify: one might say that ontological guides (or guides to ontology) concern representations (or properties of representations) that provide accurate (or approximately accurate) descriptions. I prefer Sklar's (2003) terminology of "ontological guides" because descriptions are linguistic entities, while not all representations (or properties of representations) are linguistic entities. Similarly, we could also talk about ontological guides as representations that resemble the object being represented in some manner. But, again, it is not necessarily the case that accurate representations resemble the object being represented, as is clear from accurate (linguistic) descriptions. This is why the somewhat amorphous terminology of "ontological guides" is better suited for my purposes.

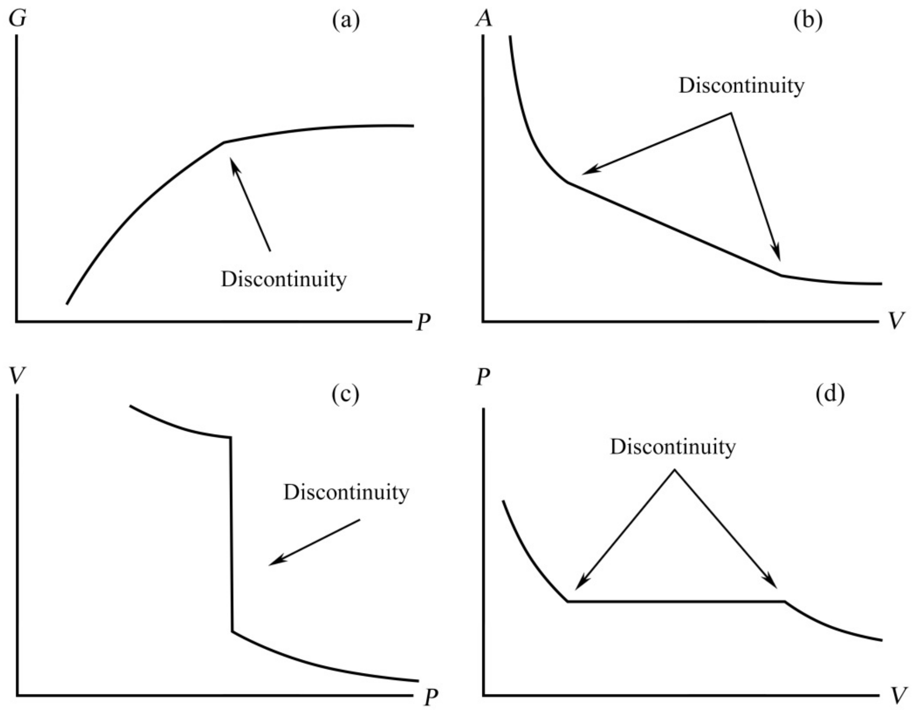


Figure. 3.1: Graphs displaying a first-order phase transition. Graph (a) displays the Gibbs free energy (or Gibbs thermodynamic potential) G as a function of the pressure P, graph (b) displays the Helmholtz free energy (or Helmholtz thermodynamic potential) A as a function of the volume V. Graphs (c) and (d) display functional relations between P and V. Based on Stanley (1971, 31).

Accordingly, although the various discontinuities appearing in the graphs represent phase transitions in methodological and epistemological senses, they are not *guides to ontology* in the sense that we can validly and soundly infer from the scientific representation of phase transitions that such phenomena arise solely in infinite systems. Thus, we *need* to be able to distinguish between a representation that (allows for sound inferences to be extracted but at the same time) misrepresents some system of interest, on the one hand, and a representation that tells us what the system is like, on the other hand.

The take-home message from all of this is that philosophical accounts of scientific representations must satisfy certain requirements. In particular, any tenable theory of scientific representation must make room for three notions. The first is that of a representational *artifact*, i.e., we must be able to identify whether or not inferences extracted from a representation are licensed. The second and third notions concern *misrepresentation* and *ontological guide*, respectively.[34] In other words, given that a representational structure allows for sound inferences to be extracted, we still must be able to make a distinction between representations that misrepresent the object being represented and those that act as guides to ontology that tell us what the object is like. Recall, part of what I will argue in this chapter is that Suárez's (2004) inferential account and Contessa's (2007) interpretational account of scientific representation are flawed, for they do not satisfy one or more of the aforementioned requirements. In the following section I continue to introduce terminology and concepts from the philosophical literature on both scientific and non-scientific representation, which will be pertinent for arguments to come.

[34] The idea that a tenable theory of representation must allow for misrepresentation is identified also in Stich and Warfield (1994, 6-7) in the context of mental representation and by Frigg (2006, 51), Suárez's (2003) (and also implicitly by Hughes (1997)) in the context of scientific representation.

3.3 REPRESENTATION

In a typical case of representation, scientific or otherwise, there are two objects that play conspicuous roles: the vehicle and the target.[35] The classical example is one of a map, say, of a city. The map is the vehicle of representation, and the city is the target. One of the important features of this type of representational relation is that a competent agent can use the map to reason and make inferences about the city. Following Swoyer (1991) and Contessa (2007), we call this process *surrogative reasoning*, or *surrogative inference*, and dub these types of representations *epistemic representations*.[36] It is worthy to note that we allow surrogative reasoning, in principle, to be any kind of reasoning—deductive, inductive, abductive, etc. Contessa (2007) further differentiates between valid and sound surrogative inferences. The former includes licensed inferences from vehicle to target, while the latter—called "faithful representations"—pertains to inferences that are valid and also "true of the target" (51).[37] Such concepts are important for they are meant to ground a robust notion of representation that satisfies the requirements discussed in Section 3.2. In particular, invalid inferences correspond to representational artifacts, valid but unsound inferences (leading to false conclusions about a

[35] It is also common to call the vehicle the *source* and the target the *object* (Suárez 2003, 2004).

[36] Epistemic representations have been called cognitive representations by Suárez (2004) and structural representations by Syower (1991). I will adopt Contessa's (2007) terminology because I think it is most apt.

[37] The distinction can be motivated with the example of using old and outdated, versus new and updated, maps. Both allow for valid surrogative inferences about the terrain, but the new and fully updated map will allow for sound surrogative inferences, while the outdated map will give rise to inferences that are not sound (although some inferences may still be sound). Note that Suárez (2003, 229) takes the "means of representation" to be the relation between vehicle and target that must obtain in order to allow for surrogative reasoning. However, given the distinction between valid and sound surrogative inferences, it should be clear that terminology of "means" is too coarse-grained for a proper investigation. This is so because the means allowing for valid reasoning, and those allowing for sound reasoning, might be very different.

target) correspond to misrepresentations, and sound inferences correspond to faithful representations that might ultimately act as ontological guides.[38]

In contrast, I will argue in the following sections that the above framework does not allow for important notions such as representational artifacts and misrepresentation, and is thus deficient. Instead, I suggest that we think of epistemic representations as *intentional* objects that come with *reference*, *semantic contents* and a *representational code*. I submit that it is in virtue of their intentional (semantic) contents that epistemic representations can be used for surrogative reasoning. The purpose of the rest of this section is to introduce and elaborate on the concepts of intentionality, contents, reference, and code, by making a connection between discussions of mental, linguistic and pictorial representations, on the one hand, and scientific representation, on the other.

3.3.1 Intentionality, Reference, Contents and Code

The word "intentionality," as it arises in the literature on the philosophy of mind—not to be confused with "intention" as in having a purpose or end—derives from the Latin verb *intendere*, which means directed towards something (Jacob 2010). It is used to convey the directedness or aboutness associated with mental states with respect to objects and state of affairs in the world (Siewert 2006). However, although non-mental intentionality might very well derive from the mental, it seems that representations such as pictorial, linguistic and scientific representations all

[38] There is a potential source of confusion here. According to Contessa (2007), "faithful representations" are representations that allow for sound inferences, i.e., true conclusions, to be made about the target. However, the question of whether or not faithful representations provide accurate descriptions that can be used as ontological guides is ignored. This is why I state that faithful representations "might" act as ontological guides. Section 3.6 elaborates on this issue. There I will argue that we must make a distinction between representations that provide sound inferences, and those that can also act as ontological guides.

exhibit the phenomenon of intentionality in the following sense: A vehicle is a representation of a target only if it is directed toward or about a target (in some sense).[39]

Moreover, we can further extend the connection between mental, linguistic and pictorial representations, on the one hand, and scientific representations, on the other. Mental, linguistic, and pictorial representations all come with *contents*. For example, my fear that "Kurt Cobain was murdered" contains the (propositional) content that, in fact, Kurt Cobain was murdered. Similarly:

> For a person to believe that the flower is in the vase is for the person to stand in the relation constitutive of believing to a state of the person, a mental representation, which *means* that the flower is in the vase. (Warfield and Stich 1994, 3).

"Means" here is synonymous with representational content. Similarly, consider examples from linguistic and pictorial representations. The sentence "Superman is Clark Kent" contains the (propositional) content that Superman is Clark Kent, while a painting of Superman flying will contain the pictorial content that Superman is flying. In both cases, it is the representational content that ascribes properties to Superman (e.g., the ability to fly) and conveys information about him (e.g., that he is Clark Kent). And so it is in virtue of contents that one can use the vehicle to make inferences and extract information about the target.

In addition, in all such representations the vehicle *refers* to the target of representation. For example, the proper name "Superman" and definite description "the flying superhero with

---

[39] This point is not meant to be controversial. I take is as evident that the concept of representation presupposes that something is being represented, and it is in this minimal sense that representations are intentional objects.

blue tights and a red cape" both refer to Clark Kent.[40] What is noteworthy is that on most accounts of, say, linguistic representation, it is the contents (along with context and circumstances of utterance) that ground reference (Speaks 2014). Similarly, it is the contents of a picture (or mental state) of Superman that determines a relation of reference from the picture (or mental state) to Superman. The contents are fundamental, while reference (or "denotation") is derivative.

Last is the notion of a representational code, which licenses valid inferences from vehicle to target. One may think of a code as a function from a vehicle to its contents. For example, a simple look at a world map found in "Google Maps" will suggest that Greenland is comparable in size to Africa. This of course is not the case (Africa is about fourteen times larger than Greenland). This is an example in which a representational vehicle contains surplus features, the purpose of which is not to represent. It is important to note that Google Maps does not misrepresent the size of Greenland. Rather, the size is an artifact of the vehicle. Whether or not features of our vehicle license inferences to be made about the target is decided by what I call the code.[41] In our example, it is clear that the code is grounded via normative conventions of map

[40] I'm setting aside here worries regarding the fact that both Superman and Clark Kent are fictional. I could have equally well used an example with non-fictional persons.

[41] Other authors have discussed similar ideas. Bolinska (2013, 224) takes "informativeness" to be what I call a code: "Informativeness is the feature of a vehicle that allows a user to draw conclusions about the target system at all." However, she then continues to argue that a vehicle is informative if and only if it is constructed with the aim of faithfully representing (i.e., extracting sound inferences about) a target. But this cannot be the case. This should be clear from political caricatures in which the representational vehicle constructed aims at *misrepresenting* the states of affairs. That is to say, in such caricatures the idea is that the viewer will understand that the representation attributes qualities to its target which the target does not have. See Section 3.5 for more on the issue.

Contessa (2007, 55) talks about "scope of representation," which, on his account, is grounded in an "interpretation." His account will be discussed in Section 3.4-3.5. Giere (2004, 748) and Teller (2001, 401) at times use the word "relevance" for code. But "relevance" is also used as a criterion for constructing and/or choosing a particular representational vehicle to begin with, so I will not use "relevance" in this way. Instead "code," as in the "code of representation," is especially apt to capture the relevant notion of "scope" because of its double meaning. On the one hand, a code, understood as a key, legend or guide, is needed in order to make use of a representational vehicle for surrogative reasoning. On the other hand, understood as a cryptogram or cipher, the code of a representation is not always known and so it must be "deciphered," so to speak.

making. But there are cases, such as tree rings representing the age of a tree, in which the code must be deciphered through empirical investigation. There are also cases, such as when deciding which operators in quantum mechanics will count as genuine observables and selecting superselection rules, in which the code must be deciphered through theoretical investigation.

To end, a study of the literature on mental, linguistic and pictorial representations—only hinted at in this section—suggests that non-scientific representations are intentional objects that come with reference, code and contents. Since scientific representations are first and foremost *representations*, I propose we think about them in the same manner. In the following three sections I will illustrate how canonical accounts of scientific representation are flawed—i.e., they do not satisfy the requirements mentioned in Section 3.2—because they lack reference to the important features discussed here such as code and contents.

3.4 THE INFERENTIAL ACCOUNT AND REPRESENTATIONAL CODE

Suárez's (2004) inferential conception of scientific representation may be formulated as follows:

> [Inf.] A represents B only if (i) the representational force of A points towards B, and (ii) A allows competent and informed agents to draw specific inferences regarding B. (Suárez 2004, 773)

According to Suárez, "denotation takes care of [representational] force," so the purpose of condition (i) is to ground the idea that a vehicle must *refer* to its target.[42] Condition (ii) relates to

[42] First, in order to alleviate potential confusion, note that "denotation" is sometimes used restrictively between linguistic entities such as terms and non-linguistic entities such as concrete objects. Here it is used as a labeling or referential relation between any object, linguistic or otherwise, and any other object.

Second, for the interested reader, I think subject of force merits some additional discussion. Specifically, Suárez (2004, 768; 771) takes *representational force*, or "force" for short, to be an irreducible feature of representation, its "essential directionality," which acts as a necessary condition for something to be a representation. However, what exactly *is* the representational force is more delicate manner. Suárez (2004) describes it as follows:

> This is the capacity of a source to lead a competent and informed user to a consideration of the target. Force is a relational and contextual property of the source, fixed and maintained in part by the intended representational uses of the source on the part of agents: No object or system may be said to possess representational force in absence of any such uses… The source's force varies with intended use. In each case an informed and competent agent will be led, upon considering the source, towards the correct target… (768-9)

> The establishing and maintaining of representational force … requires some agent's intended uses to be in place; and these will be driven by pragmatic considerations. (773)

Several remarks are in order. To begin, as mentioned, insofar as representational force is established and maintained by agents and their intended use, it would seem that all that is meant by force is conventionally stipulated denotation. On the other hand, the claim that an "agent will be led, upon considering the source, towards the correct target…" (768-9) makes it seem like representational force is related to an agent's ability to produce valid or sound surrogative inference. What else can "correct" mean? However, Suárez (2004) clearly divorces between the notions of representational force and surrogative inference, and sees both as necessary, but independent, conditions on scientific representations.

In addition, how is it that force is contextual, given that it is fixed by an agent? Presumably what is meant it that a particular context guides the fixing of force. But then it would seem that a new context allows for the same force to be "re-fixed." The claim that stipulated denotation changes from context to context based on the whim of agents, hides a deep philosophical issue: the phenomenon of reference change in the context of direct reference. This is a well-known problem with Kripkean causal-direct theories of reference (Reimer 2009). I will not belabor on the issue here. Instead, the point is that, when discussing the issue of representation, one must further elaborate on how force can be contextual. For example, it is clear how change of reference is possible if reference is cashed out via descriptions, but this is not what authors (e.g., Contessa 2007, Goodman 1976, Hughes 1997) mean when they talk about force as denotation in the context of scientific representation. All this is to say that we should be suspicious of Suárez characterization of force, which, by his own account, should be first and foremost understood as inherent directionality.

Last, one may wonder whether denotation characterized as a standing-in-for relation is even sufficient for directionality. Directionality is an anti-symmetric relation like representation. It is usually agreed that denotation is also an anti-symmetric relation: my name denotes me, but I do not denote it. One the other hand, it would seem that a standard example of denotation qua standing-in-for is the phenomenon of change of variables from mathematics which is a symmetric relation. For instance, if we want to find roots of sixth order polynomial, $x^6 + 9x^3 = 8$, we can let the letter u label, "stand in for," or denote x^3: $u \equiv x^3$. The new equation reduces to quadratic equation $u^2 + 9u = 8$ that is easily solved with the roots $u = 1, 8$ corresponding to the roots $x = 1, 2$. This is a classic case of denotation qua standing-in-for. That said, notice that there is nothing about u which directs itself to x^3. u is not about x^3, it is just a place holder. Moreover, we are free at any point in time in the calculation to switch between u and x^3, just as a person who has two names (which they equally embrace) can switch between the two. Thus, denotation—again, understood as "standing for," or labeling, or identification—is a *symmetric* relation. Accordingly, denotation cannot capture the essential directionality of representation.

how representations can be used as surrogates to reason about a target. However, as is made clear by Contessa (2007, 61) and emphasized by Bolinska (2013, 225-226), Suárez is quiet on *how* exactly users are able to make inferences about a target by using a representational vehicle. Moreover, the inferential conception leaves the phenomenon of misrepresentation unaccounted for. I would like to add to this critique in order to stress my claim that representations come with a code. To that effect, consider the following passage from Suárez (2004, 772; emphasis mine):

> Suppose that I stipulate that the paper upon which I am writing represents the sea, and the two pens that I use to write represent ships on the sea. *This act of denotation allows us to correctly draw* a few *inferences* about the ship-on-sea system on the basis of a consideration of the pens-on-paper system, such as, for instance, that the trajectories of ships may cross and that they may crash. I may have just as well stipulated that the pens will represent the sea and the paper will represent the ships; but this correlation seems counterintuitive and unnatural. I would argue that it seems so because it is less *informative,* since the relative movements of pens and paper [cannot] allow us, for instance, to infer the possibility that the two ships may crash. The ships-on-sea system is more objectively characterized by the first denotational arrangement than by the second.

First, I take issue with the claim that just by stipulating that the paper denotes the sea, and the two pens denote ships we can then "correctly" draw inferences about the ship-on-sea system. How, for instance, may we correctly infer by inspecting the paper and pens that "the trajectories of ships may cross and that they may crash?" The answer is that we cannot because there is no representational code licensing any inferences from the vehicle of representation to the target. To see this, imagine that instead we stipulate that the paper denotes the color blue and the pens denote the numbers three and four. Are we then to infer "on the bases of consideration of the pens-on-paper system" that the trajectories of the numbers three and four may cross or crash as they traverse on top of the color blue? Or, to use the original example, why didn't Suárez infer that the sea has the same texture of paper that denotes it and that ships are excellent writing instruments? The reason is that Suárez, as well as all competent agents submersed in symbolic and linguistic norms, implicitly assume that "denotes" brings with it a representational code.[43] But here's the rub: denotation is a labeling procedure and so it does not come with a set of inference rules or norms, i.e., with a code. It is the code that tells us which inferences are valid and which are not. In many contexts, the code is implicit. For instance, we know that although a map represents structural relations between streets in a city, the texture of the map does not represent anything at all. However, in some cases, such as in my example with the numbers three, four, and the color blue, a code must be deciphered in order to license any inferences from vehicle to target.

Certainly, there are numerous occasions in the history of science in which some representational structure, say, the square root of negative one, is discarded as a representational artifact, only to be later rediscovered as a genuine representational structure, an imaginary number. For instance, whereas Einstein abandoned the addition of the cosmological constant (Λ) to his gravitational field equations after Hubble's observation that the universe is not static, recent inflationary cosmologies have readopted Λ as the mathematical structure representing the

[43] Note that it will not do to try to sneak in the concept of a code through the language of "competent and informed user" (Suárez 2004, 773). It is reasonable to demand that an account of scientific representation *per se* will say more about how competent and informed users are licensed to make certain inferences and not others. See Contessa (2007, 61) and Bolinska (2013, 225-226) for a similar critique of Suárez.

exponentially-fast expansion of the early universe (Earman 2003a, 559-560). Similarly, previously discarded negative energy solutions were taken by Dirac to represent particles with positive charge and energy (and consequently predicted the positron) (Pashby 2012).

This leads us to a second qualm that I have with Suárez's claims in the above quote. In particular, he says that choosing a different vehicle, in which the pens denote the sea and the paper denotes the ships, we arrive at a "less informative" representation. However, *pace* Suárez, without a representational code in place both representational systems—arrived at through mere stipulated denotation—are equally uninformative. Let us be clear here on what "denotation" is, lest we place too much philosophical burden on the notion. Denotation is synonymous with (or taken to be a type of) *reference*, which is the primitive "standing for" or labeling relation (Goodman 1976, Chapter 1). Paradigmatic examples include proper names. To say that a pen represents a ship by the mere act of stipulated denotation is to say that the pen labels or stands in for a ship. However, it does not follow that *any of the properties or relations that apply to the pen also apply to the ship*. This point is analogous to John Norton's claim that there is no "principle of similarity" that states "that things that share some properties must share others" (2011, 2). Only I wish to take things a step further by emphasizing that denotation alone does not even ground that things share the first set of properties to begin with, let alone any other ones. It should be clear then that mere denotation, without a representational code, does not license any inferences to be made from vehicle to target. Accordingly, Suárez's inferential account satisfies none of the requirements needed for a tenable theory of scientific representation.

One might object, however, that since Suárez's (2004) account is deflationary-in-spirit, my criticism can be avoided. After all, Suárez claims that "representational force" (roughly, a directedness from vehicle to target) and surrogative reasoning are aspects of scientific representation that are necessary, not sufficient, for representation (773). It might very well be that one must also adopt or decipher code, but this will be done on a case-by-case basis. In reply I'd like to make two points. First, as stated above, it is not my intention here to produce a knock-down argument against the inferential account. My goals are to illustrate the need for a representational code that licenses valid inferences from a vehicle to target, and to emphasize that the mere act of denotation will not do the trick. These goals have been met. Second, it seems reasonable to demand that an account of scientific representation will say more about how competent and informed users can use vehicles for surrogative reasoning, and how a vehicle gains its representational force.[44] To that effect we can remain deflationary-in-spirit, stating that a detailed account of content determination and code decipherment might necessitate a case-by-case study, but also be explicit about the need for a representational code and contents. Accordingly, we might reformulate the inferential account as follows:

> [Inf.*] A represents B only if (i) the representational contents of A determine that A refers to B, and (ii) A allows competent and informed agents to draw specific inferences regarding B through a deciphered representational code.

[44] See footnotes 42 and 43. I would add that Suárez's (2010) presentation of his own (2004) inferential account amends for some of the faults identified here. For example, Suárez (2010, 98) takes a different line on "representational force," emphasizing that it is established by the norms that govern scientific practice, instead of mere convention or stipulation. In this context, the inferential account to scientific representations is more analogous to inferential accounts found in the philosophy of language literature (e.g., Brandom 1994). Clearly, it is beyond the scope of this dissertation to treat the details of this amended inferential account of scientific representation. However, I will say that from my perspective such inferential approaches, especially in a scientific context, seem like a case of putting the cart before the horse: It is in virtue of the fact that a representational vehicle represents a target that the former can be used to make inferences about the later, not the other way around. That said, a true inferentialist would not be moved by such intuitions, and it would be interesting to see if a thorough and developed inferential account (such as Brandom's (1994)) can be extended to scientific case studies.

It is worthwhile here to connect back to our motivating example from Section 3.2 (and Chapter 2) of phase transitions. The inferential account tells us that discontinuous partition functions (or thermodynamic potentials)[45] represent concrete phase transitions because the former both refer to (i.e., the representational force is directed at) and can be used to make inferences about the latter. However, it is silent on how such discontinuities refer to phase transitions and can be used for surrogative reasoning, especially in light of the fact that there are no physical discontinuities in the concrete systems undergoing phase transitions. Moreover, as I have argued, without a deciphered representational code, it is not possible to make any inferences from the discontinuities to the phase transitions. In the following section I will emphasize the role of contents in representation and the importance of misrepresentation.

3.5 THE INTERPRETATIONAL ACCOUNT AND REPRESENTATIONAL CONTENTS

The notion of a representational code is similar to Contessa's (2007) notion of interpretation in the sense that they both license inferences from vehicle to target. However, Contessa's account, which comes in general and specific versions, is deficient. The general version, which states that an interpretation takes place when a user takes facts about the vehicle to stand in for (i.e., denote) facts about the target, sheds little light on why such a process can license inferences from vehicle to target. This is the same criticism I put forth against Suárez in the last section. For instance, I can take this sentence to denote the mass of the electron and the sentence that came before it to

[45] That is to say, the kinks and jumps in graphs a-d in Figure 3.1.

denote the charge of the electron, but it does not seem to follow that I can make any valid inferences about the electron or its charge from this interpretation. We turn then to the specific version, which Contessa calls an "analytic interpretation" (57). I will argue that an analytic interpretation does not allow for misrepresentation so that it does not satisfy the requirements set on theories of scientific representation.[46] My argument comes in two parts, for there are two different senses in which misrepresentation is missing from the interpretational account. First, I will suggest that on an analytic interpretation all valid inferences will also be sound. If I'm correct about this, it will also be clear that analytic interpretation does not capture the logic of surrogative reasoning—as will be illustrated with an example from diagrammatic representation below—thereby adding another reason to be skeptical of the account. Second, *even if* the interpretational account is sensitive to the distinction between valid and sound inferences, it is still insensitive to the distinction between misrepresentations and ontological guides (accurate representations) when both structures allow for sound inferences.

The following two subsections will be dedicated to presenting Contessa's position and illustrating its shortcomings, respectively, via the first sense of misrepresentation. Limitations of the interpretational account concerning the second sense of misrepresentation will be presented in Section 3.6. The upshot is that the distinction between valid and sound inferences, and the distinction between misrepresentation and ontological guides, cannot be gained without appealing to representational content.

[46] To emphasize: such criticism ought to be worrisome, for the "key advantage" of Contessa's account is that "it renders the concept of epistemic representation applicable not only to instances of truthful or accurate representation, but to those of misrepresentation as well," as is identified, for instance, by Bolinska (2013, 222).

3.5.1 Contessa's Analytical Interpretation

To begin, it is necessary to explicate Contessa's characterization of an analytic interpretation in detail, and so I quote him in verbatim (2007, 57-58):[47]

> An analytic interpretation of a vehicle in terms of the target identifies a (nonempty) set of relevant objects in the vehicle ($\Omega^V = \{o_1^V, \ldots o_n^V\}$) and a (nonempty) set of relevant objects in the target ($\Omega^T = \{o_1^T, \ldots o_n^T\}$), a (possible empty) set of relevant properties of and relations among the objects in the vehicle ($P^V = \{nR_1^V, \ldots, nR_m^V\}$, where nR denotes an n-ary relation and properties are construed as 1-ary relations) and a set of relevant properties and relations among the object in the target ($P^T = \{nR_1^T, \ldots, nR_m^T\}$), …
>
> A user adopts an analytic interpretation of a vehicle in terms of a target if and only if
>
> 1. The user takes the vehicle to denote the target.
> 2. The user takes every object in Ω^V to denote one and only one object in Ω^T and every object in Ω^T to be denoted by one and only one object in Ω^V.
> 3. The user takes every n-ary relation in P^V to denote one and only one relevant n-ary in P^T and every n-ary relation in P^T to be denoted by one and only one n-ary relation in P^V.

Furthermore, the above account is meant to underlie the following set of inference rules (61):

[47] For simplicity, I have extracted Contessa's (2007) talk of functions, and this includes his "Rule 3" (57-58, 61-62). My claims can be easily extended from objects, properties, and relations to include functions as well.

Rule 1: If o_i^V denotes o_i^T according to the interpretation adopted by the user, it is valid for the user to infer that o_i^T is in the target if and only if o_i^V is in the vehicle,

Rule 2: If o_i^V denotes o_i^T, . . ., o_n^V denotes o_n^T, and nR_k^V denotes nR_k^T according to the interpretation adopted by the user, it is valid for the user to infer that relation nR_k^T holds among o_i^T, . . ., o_n^T if and only if nR_k^V holds among o_i^V, . . ., o_n^V, …

Last, but importantly, the concept of validity is cashed out as follows (p. 62):

> If a user adopts analytic interpretation of the vehicle, then an inference from the vehicle to target is *valid* (for that user according to that interpretation) if and only if it is in accordance with Rule 1 [and/or] Rule 2 …

That is to say, the interpretational account is superior to the inferential account in the sense that it does distinguish between valid and invalid inferences, and so it can accommodate the notions of artifact and code.[48]

---

[48] It is worthwhile to word a worry regarding a possible circularity in Contessa's account. The qualification "relevant" throughout Contessa's discussion is a serious fault of the account. This is so because in order to adopt an interpretation, one must know what the "relevant" properties are. However, in order to know what the relevant properties are, it seems one would need to first adopt an interpretation. As an example consider again a map. It is valid to infer that if two streets cross in the map they also do so in the city, but it is not valid to infer that if the streets are several inches long in the map then they are also several inches long in the city. However, before adopting a representational code, which is what an analytic interpretation is meant to ground, how is one to know that streets crossing in the map are relevant (while street lengths are not relevant) properties to adopt for the purposes of valid inferences? The point is that the qualification "relevant" might make Contessa's account circular: one needs to adopt an analytic interpretation (or code) in order to know what are the "relevant" objects and properties in the vehicle and target, but in order to adopt an analytic interpretation (or code) one needs to first know what are the "relevant" objects and properties. Accordingly, Contessa ought not to let too much depend on such a qualification.

Another noteworthy worry is the fact that if we flesh out the word "denotation" by "a mapping from objects, properties, and relations in the vehicle to the target," then adopting an analytic interpretation reduces to identifying an isomorphic relation between vehicle and target. It is questionable, then, whether the interpretational

3.5.2 The Need for Representational Contents

I will argue that on an analytic interpretation all valid inferences are also sound, so that the account leaves no room for misrepresentation in this sense. Moreover, the logic of surrogative reasoning, in which an agent first investigates a vehicle, and only later extracts inferences about the target, is not accounted for. My purpose is to stress the need for representational content. I will be considering two examples that illustrate my claims. The first focuses on diagrammatic representation, while the second focuses on both pictorial and mental representation.

Consider the following vehicle and target of a diagrammatic representation (see Figure 3.2):

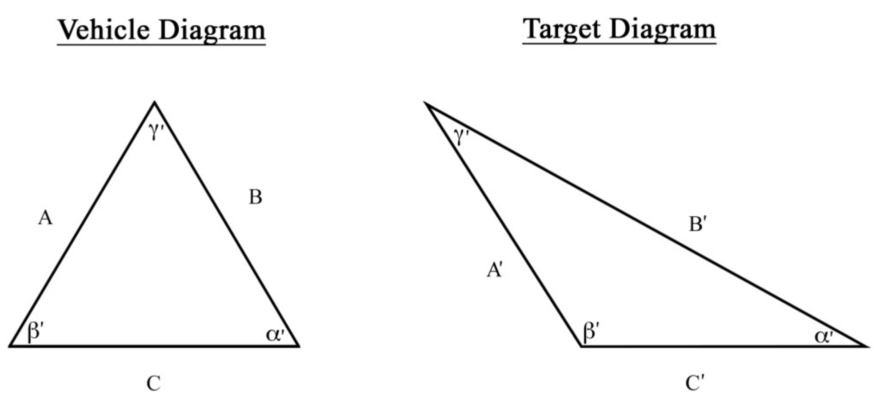


Figure 3.2: An equilateral triangle (the vehicle) and an obtuse triangle (the target).

The vehicle diagram is of an equilateral triangle with three equal sides A, B and C, with corresponding angles $\alpha, \beta,$ and γ. The target diagram is an obtuse triangle with unequal sides A',

account's claim to novelty can withstand scrutiny when compared to, say, Swoyer's (1991) structural account of scientific representation.

B' and C', with corresponding angles α', β', and γ'. Let us adopt an analytic interpretation of the vehicle in terms of the target such that A denotes A', B denotes B', and so on. Given the fact that triangle ABC is a triangle with equal sides we can deduce the following property belonging to triangle ABC (or "relation" among the angles): $\alpha = \beta = \gamma$. That is to say, after investigating the vehicle, we conclude that the angles interior to triangle ABC are all equal, and the question is what valid inferences we may conclude about our target using our vehicle. One might think that a simple substitution will avail that $\alpha' = \beta' = \gamma'$, which would turn out to be a valid but unsound inference that misrepresents the target. However, this is not the case because by adopting an analytic interpretation we have to take the relation $\alpha = \beta = \gamma$ to denote some relation *in the target*. The relation $\alpha' = \beta' = \gamma'$ is not one that appears in the target, and so we cannot let $\alpha = \beta = \gamma$ denote $\alpha' = \beta' = \gamma'$ for two reasons.

First, Contessa's own account does not allow for such identification for he claims that the "set of relevant properties and relations among the object" must be "in the target" (p. 57).[49] Second, it is not possible for a relation in the vehicle to denote a relation in the target that doesn't exist. Instead, possible denotable relations might include $B' > C' > A'$ or $\beta > \gamma > \alpha$. Accordingly, we'll be able to validly infer from the fact that $\alpha = \beta = \gamma$, that $B' > C' > A'$ or $\beta > \gamma > \alpha$. This is problematic, again, for two reasons. First, as mentioned before, analytic interpretation leaves no room for misrepresentation in the sense that some inferences are valid but not sound. Second, the account distorts the logic behind surrogative reasoning (or diagrammatic reasoning in our case) in which a user can extract inferences about a target by investigating the vehicle. I'll elaborate on this second point below, but first a word in reply to a potential objection.

---

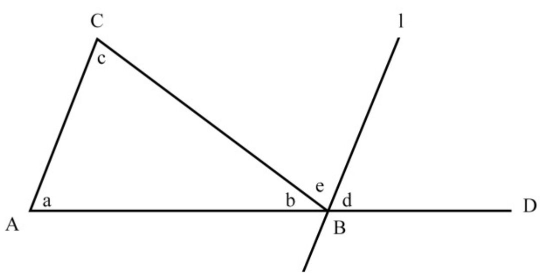
[49] Also see Rule 2 above: the "relation nR_k^T holds among o_i^T, . . ., o_n^T" where o_i^T, . . ., o_n^T are objects *in the target* (61).

The reader might worry that it may be possible to give a more charitable reading of Contessa's account, or amend it in an appropriate way, so that we may make a distinction between valid inferences that are sound and those that are not sound. Given the type of logical framework used by Contessa to present his analytic interpretation, so the complaint goes, surely we can speak of, say, stipulated relations that are syntactically true of some target but not semantically true of the particular target at hand. In other words, we allow a relation in the vehicle to denote a relation that is "in the target" in the sense that we stipulate that such a relation exists. For instance, considering Fig. 3.2, the relation $\alpha' = \beta' = \gamma'$ might not be in the target in the sense that is not semantically true of the target, but we may still stipulate that the general relation equality-between-two-objects is in the target, and adopt an analytic interpretation to infer via Rule1 and Rule 2 that indeed $\alpha' = \beta' = \gamma'$. This will be a valid inference that is not sound, thereby allegedly allowing for a notion of misrepresentation. However, this is not the case. If indeed such an inference procedure is possible then by the same account we could adopt an interpretation of a target in terms of a vehicle in which both are identical, and still extract inferences that are valid but not sound.[50] But if the vehicle and target are identical, then a valid but unsound inference is not a mark of misrepresentation and, again, the interpretational account is insensitive to misrepresentations.

Moreover, to return to the second point from above, the interpretational accounts gets the logic of surrogative reasoning wrong. To see this, consider a similar diagrammatic representation from Euclid's *Elements* (proposition 32) (see Figure 3.3):

[50] For example, let the target of Figure 3.5.1 be both the vehicle and the target of representation and adopt an analytic interpretation in which A' in the vehicle denotes A' in the target, B' in the vehicle denotes B' in the target, and so on. Also let the greater than relation $>$ denote the equality relation $=$. It will follow from the fact that $\beta > \gamma > \alpha$ in the vehicle, that $\beta = \gamma = \alpha$ in the target. This is a valid inference that is not sound. But since the vehicle and the target are identical this is not a case of misrepresentation. How can a vehicle misrepresent a target if the two are identical? In other words, on this reading (of Contessa's (2007) account) valid but unsound inferences do not correspond to misrepresentations.

Vehicle Diagram



Target Diagram

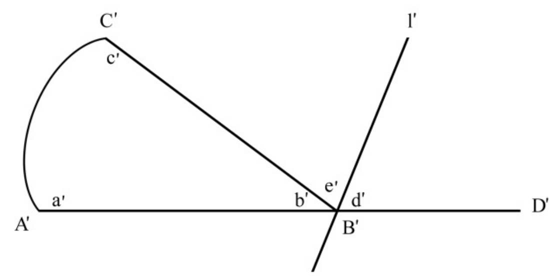


Figure: 3.3 A triangle ABC with a line l parallel to side AC (the vehicle), and an almost identical figure in with the exception of side AC bent outward as an arc.

One can use the vehicle diagram to demonstrate that the interior angles in the triangle ABC are equal to two right angles via steps 1-4 below.[51] We can further adopt an analytic interpretation in which the vehicle diagram denotes the target diagram, and A denotes A, B denotes B, and so on, with a denoting a' and c denoting c'. Similarly, we can let relations 1-4 within the vehicle denote relations 1'- 4' within the target (i.e., $c = e$ denotes $c' > e$, etc.):

1. $c = e$ (alternate interior angles)
2. $a = d$ (by prop. 29)
3. $b + e + d =$two right angles (by prop. 13)
4. Thus, $a + b + c =$two right angles

1'. $c' > e$ (by construction)

2'. $a' > d$ (by construction)

3'. $b + e + d =$two right angles (by prop. 13)

4'. Thus, $a' + b + c' >$ two right angles

[51] See Burton (2011, 156), from which this example is taken, for a proof.

It will follow then that by working through the steps 1-4 of Euclid's proof for proposition 32, we can demonstrate that $a' + b + c' >$ two right angles. But this is absurd. The logic of surrogative reasoning (in the context of Euclidean proofs with diagrammatic representations) allows us to reach conclusions about a target through the investigation of a vehicle, *in virtue of the fact that vehicle is about the target in an appropriate way*. Instead, what we had to do is investigate *both* the target and the vehicle and then define various denotational relations. Recall, proposition 32 is not meant to be about the one triangle that appears in the proof but about all triangles. It is because triangle ABC "represents" any generic triangle that we can infer that all such triangles have interior angles equal to two right angles. If triangle ABC, along with its properties and relations, were to solely "denote" generic triangles *a la* analytic interpretation, then we would not be able to make such an inference any more than we can infer that the internal angles of the object in the target diagram (in Fig. 3.3) are greater than two right angles ($a' + b + c' =$ two right angles).

In order words, *contra* Contessa (2007, 66) it is *not* "*in virtue of* their interpretation of the vehicle in terms of the target that users would be able to perform surrogative inferences from vehicle to the target."[52] Rather, users can perform surrogative reasoning in virtue of taking the vehicle and its components to both refer to (denote) the target, but also be *about* the target in a manner determined by the representational *contents* of the vehicle (as was made clear by the discussion of non-scientific representation in Section 3.3).

[52] Recall that for Contessa "interpretation" is a term of art. Colloquially, we might say that it is in virtue of an interpretation that a user is able to perform surrogative inferences from vehicle to target *so long as the interpretation is determined by the representational contents of the vehicle*. See Bolinska (2013, 227-228) who also complains about Contessa's odd use of the term and appeals to the "ordinary" notion of interpretation to capture the concept of "informativeness" (which for her is a term of art).

Consider a second example of a pictorial representation of a white cube (Figure 3.4) and the mental representation associated with perceiving the pictorial representation below.

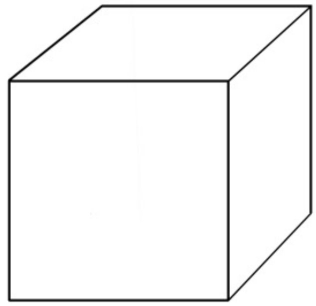


Figure 3.4. A white cube.

How do we know that this is indeed a pictorial representation of a white cube, and not, say, of a green sphere? How do we know that our perceptual mental representation is of a white cube and not a green sphere? According to Contessa (2007), the above pictorial representation can be either, depending on which interpretation one chooses. Certainly, as intelligent and purposive agents we can let cubical objects denote spherical ones, and let whiteness denote greenness. Such a stipulation will allow agents to communicate about green spheres by appealing to white cubes. However, the reason we know that the pictorial representation is about a white cube is because we *see* a white cube in the picture.[53] Similarly, our perceptual mental representation of a white cube includes the intentional *contents* of "white cube." To be clear, there is no doubt that some hidden or implicit norms are in play. Nevertheless, unless one has a well-developed theory of the normativity of determination of representational content, the most plausible view is that said representation contains the content "white cube." What determines the representational content

[53] See Hopkins (2005) and Wollheim (1987) on "seeing-in."

of scientific representations is a substantial question that merits serious philosophical inquiry—in the same manner that analogous questions regarding pictorial, linguistic and mental representations have given rise to flourishing research programs. But what is certain is that there is more than just stipulated denotation and interpretation at play.[54]

The problem is that on the analytic interpretational account, it is not possible to attribute properties and relations to the target via the vehicle. Rather, one can only *denote*. Said differently, the account only leaves room for *reference*, but not for *content*. The problem is well-known in other contexts in which representations are discussed. For instance, consider Frege's

---

[54] To emphasize this last point, consider Contessa (2007, 59-60) on the matter:

> For example, one does not need to believe that the string from which a certain pendulum hangs is massless in order to adopt an interpretation of the ideal pendulum according to which the string is massless. The knowledgeable user knows perfectly well that, since no real string is massless, the inference, though valid, is not sound…

But how is it possible to make such a "valid" inference? The ideal pendulum is an object in the ideal pendulum system—the vehicle—that denotes a massive pendulum system—the target. Call this object o_1^V. As an object in the vehicle it must denote an object in the target *a la* an analytic interpretation. Say it denotes the massive pendulum, which we label o_1^T. The ideal pendulum's zero mass is a property of an object in the vehicle (and according to the presented notation it would be $1R_1^V$, which we label by p_1^V for simplicity). Again, this property in turn must denote a property of an object in the target system, p_1^T. Let's say it denotes the massive pendulum's mass. Rule 2 then tells us that it is valid to infer that p_1^T holds for o_1^T if and only if p_1^V holds for o_1^V. And so from the fact that the ideal pendulum is massless we can validly infer, and *soundly* I might add, that the massive pendulum is massive. However, I do not see how we can ever make a valid inference that was not sound, e.g., the massive pendulum is massless. In order to make such an inference we would have to do more than just denote the massive pendulum by the massless one, or to denote the pendulums mass with zero mass.

Again consider Contessa's (2007, 62) own elaboration of his inferential rules:

> According to Rule 2, from the fact that the circle labeled "Holborn" is connected to the tab labeled "Bethnal Green" by a colored line, one can infer that a direct train service operates between Holborn and Bethnal green station.

Let us assume that there is no direct train service operating between Holborn and Bethnal Green. Is the above a valid but unsound inference? I claim the answer is "no," i.e., all reasoning that is valid will be sound, leaving no room for misrepresentation or error. The reason is that in such a scenario the colored line connecting "Holborn" and "Bethnal Green" will either not denote anything in the target and so no valid inference can be made. Or, instead, it will denote, say, the fact that the two stations are not directly connected. In such a case, a valid inference would be:

> According to Rule 2, from the fact that the circle labeled "Holborn" is connected to the tab labeled "Bethnal Green" by a colored line, one can infer that a direct train service ***does not*** operate between Holborn and Bethnal green station.

(1892/1960) famous example regarding the difference between the “the Morning Star” and “the Evening Star.” Both phrases refer to the same object; they both denote the planet Venus. But the phrases differ because they have different semantic contents. The interpretational account *a la* analytic interpretation leaves no room for the content of representational vehicles. Consequently, the account does not satisfy the requirement that a tenable theory of representation must allow for misrepresentation.

To clarify, my complaint is not meant to be a superficial one about what must denote what when an interpretation is adopted by a user for the purposes of substantiating a code that is meant to license inferences from vehicle to target. Rather, my point is that reference and denotation are notions that cannot bear the semantic weight that comes with representations. This is clear from a mere glance at other types of representations such as pictorial, linguistic and mental. The “Goodmanian” turn, which takes “denotation to be the core of representation”—a motto that is suggested by the claims made by Suárez (2004) and Contessa (2007)—must be abandoned.

3.6 MISREPRESENTATION, FAITHFUL REPRESENTATION, AND GUIDES TO ONTOLOGY

In this section I want to emphasize that there is a second sense in which the interpretational account does not capture the notion of misrepresentation. It is insensitive to the distinction between accurate representations (ontological guides) and inaccurate representations (misrepresentations) so long as both can be used to obtain sound inferences to the same degree (approximately). This is clear from the fact that Contessa dubs representations that allow for sound inferences "faithful representations," whether said representations are misrepresentations or not. Consider an illustrating example. Political caricatures clearly misrepresent various states of affairs in the world but such inaccuracies are used by a competent user to extract many sound inferences about a target vis-à-vis information communicated to the user by the caricaturist. On the interpretational account such representations are faithful for they allow for valid and sound surrogative reasoning—they are not misrepresentations. But, indeed, we know that caricatures *are* misrepresentations, and so there is a second sense (different from the first) in which the interpretational account fails to allow for misrepresentations for it is insensitive to the difference between accurate and inaccurate representations in such contexts.

Lest we stray too far from our aim of gaining insight into *scientific* representation, it would do well to attempt to connect back with examples from science. The paradox of phase transitions (discussed in Chapter 2 and Section 3.2) is a case in point. Both phase transitions (mis-)represented as discontinuities (in the partition function or thermodynamic potentials associated with a target system) and phase transitions realistically represented by sharp enough but continuous changes allow for sound surrogative reasoning. That is to say, both representational structures are (in principle) empirically adequate. However, given a background of scientific facts and theories—in this case, the atomic theory of matter—we say that one representational structure is a misrepresentation while another is not.

Similarly, scientists represent water as a collection of molecules for the purposes of investigating Brownian motion, but then they also represent water as continuous fluid when studying water flowing through pipes.[55] The problem with Contessa's notion of faithful representation is that both the water-as-a-collection of molecules and water-as-a-continuous-fluid

[55] This example is also used by Teller (2001) and Giere (2004) to stress representations can have different purposes based on an agent's (scientist) intentions.

can act as representations that are faithful (in Contessa's sense) to the same degree. Yet, the question then arises: "[Is there] no principled way of distinguishing conceptualization meant to represent how things really are with a system and these meant merely to describe 'useful fictions' about the system" (Sklar 2003, 430)? Or, similarly, is there no way to identify those misrepresentations "required to make available structures that will persist in, and underlie the empirical success of future theories?" (Ruetsche 2011, 337)

Furthermore, the problem can become even more extreme when we consider that in certain instances misrepresentations are more informative and lead to better predictions, as in the case of predicting and explaining the critical exponents associated with universality classes of systems portraying critical phenomenon (Batterman 2002, 2005). To be clear, this occurrence of superior predictions and explanations through misrepresentation is widespread and arises in varied phenomena such as the breaking of liquid drops, modeling shocks, and explaining the rainbow (Batterman 2002, 2005, 2009).[56] This means that on Contessa's (2007) terminology misrepresentations can be more faithful than ones that describe a target accurately in the sense that a misrepresentation might allow for more sound inferences to be extracted from a vehicle. But this is highly counterintuitive. How can a misrepresentation be more faithful than an admittedly (approximately) accurate representation? Accordingly, I submit that the interpretational account does not capture the notion of misrepresentation—again, it is insensitive to the differences between misrepresentations and guides to ontology that are equally well suited for extracting sound inferences—and I suggest then that we alter the terminology introduced by

[56] In order to prevent objections of the kind raised in Butterfield (2011), Callender (2001), Menon and Callender (2013), and Norton (2012) against Batterman (2002, 2005, 2009), note that I am not claiming here that such misrepresentations or idealizations are necessary. That is a claim that I will qualify and defend in Chapters 4-8. Rather, my claim is that there are many instances in which scientists aim at misrepresenting aspects of the world in order to extract sound inferences from a vehicle to a target.

Contessa (2007) to better capture our commonsensical intuitions on the matter. Let us say that while some epistemic representations allow for valid and sound inferences to be made about the target—we might call these *sound representations*—still, *faithful representations* are those which represent a target accurately. Faithful representations act as guides to ontology, i.e., they contain information about "how things really are" as Sklar (2003, 430) explains:

> [T]here is something very different between characterizing an atomic nucleus as a complex system of neutrons and protons, with these 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, and a characterization of a fissionable nucleus as a "liquid drop" held together by a "surface tension."
>
> The former description is at least a part of the structure "on the road" to our desired ultimately theory. The latter is intended, from the start, as nothing more than a weak model adequate only in the most restricted ways to characterizing what is really going on. (Sklar 2003, 439)

One might object *a la* Teller (2004, 440) that a distinction between representations qua useful fictions and representations qua ontological guides is artificial since "both are idealizations." However, this misses the point, for I'm not claiming that some representations are not idealized. After all, faithfulness and accuracy are matters of degree. Rather, I first claim that a tenable theory of scientific representation needs a conceptual and terminological framework that allows us to differentiate between those aspects of a vehicle that we think tell us what the target is like—the ontological sense of scientific representation—and those aspects that solely license sound inferences—the epistemological sense of scientific representation. Second, I submit that the relation between faithful representations and sound representations requires further scrutiny. As I see it, it is by conducting research into the determination of representational contents in the context of scientific representations that we can further our understanding of when sound representations imply faithful ones and vice versa.

Another worthwhile objection to consider may be stated as follows. In cases that I discussed, it is clear that scientists take misrepresentations to be informative. However, this does not go against either Contessa (2007) or Suárez (2004). Rather, scientists taking less informative representation to be more accurate or faithful would be a problem for their accounts. Yet, how do scientists know that the less informative representations are more faithful representations other than by means of either inference or interpretation-cum-inference? In reply, two remarks are in order. First, if the interpretational and inferential accounts are insensitive to the distinction between misrepresentations and faithful ones, then this does go against both accounts, full stop. This is so because the accounts do not satisfy the requirements set (in Section 3.2) on a tenable account of scientific representation.

Second, the cases that I am discussing (e.g., phase transitions and critical phenomena) are exactly cases in which a scientific representation is less informative (in the sense that we cannot extract many sound inferences), but also more faithful. For example, characterizing phase transition without appealing to discontinuities would allow for a more accurate representation of phase transitions. However, such characterizations may sometimes be mathematically intractable and so offer little to no information about a target system. It is well known, for example, that appealing to limiting procedures in which various variables diverge allows us to make the mathematical models representing some concrete target system more tractable so as to extract much information about such targets.[57] In the context of phase transitions, I mentioned that the misrepresentations appealed to include taking the limit in which the number of particles diverges. Nevertheless, scientists know that such representations are less faithful because of empirical observations (to the effect that boiling kettles are not infinite) and the vast empirical evidence supporting the atomic theory of matter.[58]

In conclusion, I would like to emphasize what I take to be correct about the inferential and interpretational accounts, and identify how my view diverges from Suárez (2004) and Contessa (2007). I agree with Suárez (2004) that emphasis ought to be placed on the inferential practices related to scientific representation via surrogative reasoning. However, I think more can and must be said about how scientific representations allow for surrogative reasoning. From my

---

[57] See, for instance, Batterman (2002, 2005, 2009), Butterfield (2011) and Norton (2012), and Chapters 4-8, all of which discuss such limiting procedures.

[58] A somewhat lengthy intermission might be called for at this point. It is a working assumption in the literature that phase transitions are faithfully represented by a finite-dimensional state space. I have justified this assumption by appealing to the atomic theory of matter, but one could argue that on field theoretic accounts of matter this is simply not the case. If (say) quantum field theory (QFT) is true, then matter is just excitations in quantum fields so that there is a sense in which even a boiling kettle is infinite (and this is true also for local QFT). To me this seems a bit extreme. We take phase transitions to be phenomena that are accounted for by classical statistical mechanics. We do not think about phase transitions as we do about, say, quantum non-locality. While the latter is clearly foreign to the classical world, and arises solely as a quantum effect, the former is not. Moreover, the fact that we can represent phase transitions without appealing to infinite-limit misrepresentations—e.g., as in for example, Gross and Votyakov (2000), Chomaz *et al*. (2001), and Borrmann *et al* (2000)—seems to confirm the idea that phase transitions ought to be accounted for within a classical world-picture. See Callender (2001) and Menon and Callender (2013) for a defense of the claim that phase transitions are governed by classical statistical mechanics.

As a retort, one might argue that misrepresentations analogous to the ones in the phase transitions case arise also in the context of spontaneous symmetry breaking in QFT. In such a context, it certainly is questionable whether infinite limits correspond to accurate or inaccurate representations. See Earman (2004, 191-192) who discusses the issue and is a proponent of the view in which infinite limits are not misrepresentations in the context of QFT. However, even in here, he admits, there is room for dispute (since what is the ontology of QFT is a disputed matter).

In any case, what is brought to the fore here is an interesting observation to the effect that the identification of representational artifacts versus genuine representational structures, and misrepresentations versus faithful (or accurate) representations, depends on the background theories and auxiliary assumptions that one is dealing with. While finite-dimensional state space representations of systems are faithful representations in the context of theories with a (finite) corpuscular ontology, such representations become misrepresentations when one moves to field theories. From my perspective, this point strengthens one of the general themes of this paper: deciphering a representational code and determining representational contents necessitates both empirical and *theoretical* investigation, and is part and parcel of the scientific enterprise. This point is emphasized in Section 3.7.

perspective, it is in virtue of having the semantic contents that they indeed have, that vehicles can be used as inferential surrogates for targets of representation (and not the other way around).

Next, on a general or loose conception of "interpretation," I think Contessa (2007) is correct to note that it is necessary and sufficient to adopt an interpretation of the target in terms of the vehicle in order to distinguish between valid and sound surrogative inferences. To that extent, my view expands on his by noting that "adopting an interpretation" includes deciphering a code and determining semantic contents (in addition to adopting symbolic conventions and learning the language of the particular science one is working with). However, I have argued that the specific conception of interpretation that Contessa (2007) elaborates on—the so-called analytic interpretation—is neither necessary nor sufficient for scientific representation for it would deem representations insensitive to misrepresentation and ontological guides.

We may end then by summarizing the *content-based account* of scientific representation sketched in this chapter and elaborated on in Section 3.7 as follows:

> [cont.] A vehicle *V* is an epistemic representation of a target *T* if and only if *V*'s representational (semantic) contents—determined vis-à- vis a representational code that is adopted and deciphered by intentional agents—are about T.

3.7 THE ROAD AHEAD

In this chapter I have identified requirements that must be set on any tenable account of scientific representation (Section 3.2), drawn on an analogy from non-scientific representation (Section 3.3), and put forth a critique of two canonical accounts of scientific representation (Sections 3.4-

3.6), in order to motivate the idea that we should think of scientific representations—in particular, epistemic representations that are used for surrogative reasoning—as intentional objects that come with a representational code and contents (and reference determined by contents). I argued that denotation alone can not bear the semantic weight that comes with a philosophically robust notion of representation. Consequently, I showed that a framework appealing to code and contents was needed in order to satisfy requirements set on any tenable theory of scientific representation, specifically, to allow for the notions of artifacts, misrepresentations, and faithful representations that can act as ontological guides. My hope is that such insight will pave the way for superior accounts of scientific representation, and in this section I'd like to reflect on both the main work left to be done, and on the relation between the work conducted in this chapter and the larger issue dealt with in this dissertation, namely, the EIP.

To begin, in "General Semantics," David Lewis wrote:

> I distinguish two topics: first, the description of possible languages or grammars as abstract semantic systems whereby symbols are associated with aspects of the world; and, second, the description of the psychological and sociological facts whereby a particular one of these abstract semantic systems is the one used by a person or population. Only confusion comes of mixing these two topics. (Lewis 1970, 19)

In other words, and taking into account the idea that representation are intentional objects with semantic contents, there are really two different important questions that one must answer in an investigation of scientific representation:

Q1 Content-Identification: *What are the contents of this or that representational vehicle?*

Q2 Content-Determination: *In virtue of what facts is the content of this or that representational vehicle determined?*

With respect to Q1, deciphering and adopting a representational code, which is a function from vehicle to contents, will allow for an answer. This is where emphasis on interpretation, demonstration, computation, inferential practices, etc., by functional theories of representation takes center stage. With respect to Q2, since such a question has given rise to flourishing literature in the realms of pictorial, linguistic, and mental representation, and so I submit that an analogous research program in needed in the realm of scientific representation. What is clear is that informational theories (such as those emphasizing a similarity relation) are targeting Q2, and so are not in tension with functional theories.

Thus, the central issues we are left with after my analysis concern code decipherment and content determination of scientific representations. Recall that code and contents are what allow for discrimination between representational artifacts, misrepresentations and faithful representations. We may then ask how such discrimination may be attained. Certainly, a full answer to this question cannot be given with armchair reflections. We must instead look to actual scientific theories and case studies. However, it seems to me that some general remarks can be made. In particular, that code decipherment and content determination can be ascertained by looking to science itself, and that both are part and parcel of the scientific enterprise:

> Science is replete both with schemes intended to truly characterize "how things are" and with other schemes intended only as knowingly false but useful models of the real situation. But only science itself can do the job of explicating the intended purpose of its own descriptive and explanatory schemes... there is no a priori way of deciding which descriptive schemes are to be taken as straightforward and which as useful fictive modes of characterizing the world… And much useful and sometime[s] very hard and very brilliant science is devoted to just that question. Figuring out which "fictional" schemes to apply and understanding why, in the light of the nonfictive science they can work so well, is quite often an ongoing scientific project. (Sklar 2003, 413, 438)

Accordingly, it is empirical and theoretical investigation that will allow us to discriminate between genuine representational structures and descriptive fluff, and between misrepresentations and faithful ones. For an example of the former consider how Aharonov and Bohm (1959) showed that the gauge freedom associated with the electromagnetic vector potential, which was taken to be no more than an artifact of the representational framework of classical electrodynamics, played a genuine representational role in the context of quantum mechanics leading to novel empirical predictions.[59]

An example of the latter concerns the belief, prior to Einstein's Special Theory of Relativity (STR), that the Galilean addition law was a faithful representation of the states of affairs regarding relative motion between observers: $V_{A,C} = V_{A,B} + V_{B,C}$ where $V_{X,Y}$ is to be read as "the velocity of X with respect to Y." Einstein's theoretical investigation of Maxwell's electrodynamics, along with the unsuccessful empirical attempts to discover the luminiferous

[59] This, at least, is the orthodoxy on the issue. I discuss the AB effect in detail in Chapter 6.

ether, led him to develop the relativistic addition law: $V_{A,C} = (V_{A,B} + V_{B,C})(1 + \frac{(V_{A,B})(V_{B,C})}{c^2})^{-1}$. Such a development exemplified that the original Galilean addition law was a misrepresentation of the state of affairs, while the relativistic law was a faithful one. Similarly, another empirically confirmed consequence of STR, the relativity of simultaneity,[60] along with Minkowski's geometrical reformulation of STR led to the abandonment of the representation of space and time as two independent entities—a misrepresentation—and the adoption of the view that faithfully represents space and time as a single entity, a space-time:

> The views of space and time which I wish to lay before you have sprung from the soil of experimental physics, and therein lies their strength. They are radical. Henceforth space by itself, and time by itself, are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality. (Minkowski 1908/1952, 75)

A more detailed analysis of code decipherment and content determination of scientific representations necessitates a more fine-grained study of significant scientific case studies. What should be clear is that decipherment and determination are activities dependent on empirical and theoretical investigation, constituent of the scientific method, and involve more than a sophisticated web of evermore complex stipulated denotational arrangements.

To end, I would like to reconnect with the larger issue discussed in this dissertation, namely, the EIP. I have attempted in this chapter to make headway on the part of the EIP that discusses the relation between abstract mathematical structure (the vehicle) and physical

[60] That is to say, that observers undergoing relative inertial motion will disagree on simultaneity judgments regarding spatially separated events.

phenomenon in the world (the target) by outlining, elaborating, and endorsing a content-based approach to scientific representation. The connection to idealizations—discussed more thoroughly in the following chapters to which I turn to next—comes through the idea that theoretical investigation is necessary in order to decipher a code that determines and identifies representational contents. Such theoretical investigation, as will be emphasized in Chapter 8, necessitates exploring the full and possible structure and representational capacities of a theory. But an appeal to highly idealized, abstract mathematical structure, which, at best, misrepresents the world, is essential and absolutely indispensable for such investigation. I now turn to a general discussion of the philosophical literature on idealizations.

4.0 IDEALIZATIONS IN SCIENCE

In this chapter I present a critical survey of the current state of the literature on idealizations. The main upshot will be that the sound principle is indeed the standard and most plausible justification for appealing to idealizations in science. I also identify how the pervasiveness of idealizations, abstractions and approximations in science raises various foundational problems. The punch line is that insofar as the sound principle fails for cases wherein appeals to essential idealizations are made, the problems identified reemerge with a vengeance, so to say.

4.1 INTRODUCTION

Idealizations are ubiquitous in science. Examples include, among others, the frictionless plane, the simple pendulum, point particles and test particles, nonviscous fluid flow, infinitely thin wires and infinitely long cylinders or planes, a perfect vacuum, non-interacting particles, perfectly rational agents and isolated populations. However, there is no agreed upon uniform terminology or taxonomy used either by scientist or philosophers.

This chapter has three main goals. The first is to introduce the notion of an *idealization*, and important associated notions such as *abstraction* and *approximation*, as they arise in the scholarly literature on the subject (Section 4.2). The second is to present a critical survey of the current state of the literature on idealizations (Section 4.3). The third goal is to sketch how the pervasiveness of idealizations, abstractions and approximations in science raises various foundational problems (Section 4.4).

Roughly, many such problems can be defused by appealing to some "de-idealization" or "concretization" scheme, in which more realistic scientific representations and models are accompanied by improvements in predictive output, as well as improvability of other theoretical merits such as explanatory and descriptive power. Accordingly, the main upshot of this chapter is that the most plausible and typical justification for appealing to idealizations, abstractions and approximations in science is that it is possible, in principle, to implement some sort of standard de-idealization or concretization scheme, thereby dispensing with such idealizations. Said differently, idealizations must abide by what has been come to be called Earman's sound principle (or sound principle for short):

> EARMAN'S SOUND PRINCIPLE − While idealizations are useful and, perhaps, even essential to progress in physics, a sound principle of interpretation would seem to be that no effect can be counted as a genuine physical effect if it disappears when the idealizations are removed. (Earman 2004, 191)

Or, taking into account insights presented by Laura Ruetsche (2011, 336):

> EARMAN-RUETSCHE SOUND PRINCIPLE − If a scientific account (theory, model) uses an idealization to predict or explain an effect which disappears when the idealization is removed then either the effect is an artifact of the idealization or else (if experiment confirms the effect) the theory is inadequate.

Part of the punch line of this chapter, especially Section 4.4 which discusses problems associated by appealing to idealizations, is that insofar as de-idealization schemes fail for cases of "essential" or "indispensable" idealizations, then problems associated with appeals to idealizations reemerge with a vengeance, so to speak.

In what remains of this introduction, I discuss some important terminology related to "idealizations." Namely, from a conceptual perspective, the term "idealization" is strongly associated with terms such as "abstraction," "approximation," "model," "representation," and "approximate truth." "Abstractions" and "approximations" are sometimes used interchangeably with "idealizations." "Models" and "representations" tend to be the objects that are regarded as idealized, abstracted or approximated in some sense, but some specifically discuss theories and laws as the idealized objects. "Approximate truth" is then, arguably, the relation that holds between idealized objects and the world.

However, idealizations are interesting in their own right. Accordingly, in as much as it is possible, in what follows I present conceptions and taxonomies of idealizations and do my best not to discuss in any detail the related concepts mentioned. Instead, I'll introduce necessary concepts for discussion without elaboration, and I refer the reader to the appropriate literature (also see referenced therein).[61] Scientific representations and representations in general were discussed in detail in the previous chapter. Here I'll only outline some helpful terminology so that this chapter can be read independently. The reader familiar with the terminology can jump ahead to the next section.

---

[61] For Abstractions see Cartwright (1983; 1989, Ch. 5), for a general treatment of Models see Frigg & Hartmann (2012), for Mary Hesse's approach to models and analogies in science see Koperski (2006), and for Approximate Truth see Oddie (2007).

The important ingredients of scientific representation include the following: A representational *vehicle* with representational *contents* that represent a *target* or *target system*, an *agent* (or group of agents) that is (are) undertaking in the activity of representation, and a representational *code* that allows agents to use the representational vehicle to make valid inferences about the target, the process of which is called *surrogative reasoning*. When a vehicle allows for valid surrogative inferences we call this an *epistemic representation*, and when such inferences are sound, i.e., true of the target, we call this a *sound representation*. If, further, the sound representation acts as an accurate representation, a kind of guide to ontology, which tells us what the target is like, we call this a *faithful representation*. Laws and scientific principles, along with specific agent intention and context dependent details, guide the construction of scientific representations. Models are just types of scientific representations, in fact, scientific theories, laws, and principles can also be thought of as types scientific representations.

4.2 IDEALIZATIONS, ABSTRACTIONS AND APPROXIMATIONS

The complexity of the world coupled with the limited and finite cognitive abilities of human beings, as well as intuitions of parsimony—which dictate that scientific accounts of phenomena should include only those features that are relevant for its manifestation—compel scientists to appeal to idealizations, abstractions and approximations. Such appeals concern the formulation of theories, laws, scientific models and representations, as well as the application of these objects to the concrete world. As already mentioned, the concepts of idealizations, abstractions and approximations are tightly intertwined, with no agreed upon uniform terminology or taxonomy used either by scientist or philosophers. Nevertheless, it is fair to say that, roughly, the three notions can be differentiated as follows: *Idealizations* include deliberately distorting some target system thereby creating an idealized *analogue system* that can be thought of as a false description of the original target system, but can also be in some sense approximately true. The main feature of this notion of idealization is that it introduces *simplifying falsehoods*. For example, Laymon (1995, 354) writes that the "most natural attitude to take toward idealizations … is to assume that their use introduces distortion or bias into the … analysis," Morrison (1999, 38 fn. 1) claims an idealization is "a characterization of a system or entity where its properties are deliberately distorted in a way that makes them incapable of accurately describing the physical world," and Cartwright (1989, 187) understands an idealization as a conceptual reorganization of complicated aspects of a target system with ones that "are easier to think about, or with which it is easier to calculate." As is clear from the above, some authors emphasize that idealizations are simplifications, while other concentrate on idealizations as distortions. So, for example, Frigg & Hartman (2012) characterize idealizations as "deliberate simplification of something complicated," and continue to call idealizations that involve distortions, "Galilean Idealizations," while those that involve "stripping away" details, "Abstractions" (or "Aristotelian Idealizations"). Other authors, e.g. Jones (2005), Weisberg (2007a, 2013), and Lind (1993) take idealizations to be distortions, allowing for some to be simplifications while others not. Shaffer (2012, 19-20), on the other hand, insists that idealizations must be accompanied by simplifications. He argues that the process of merely distorting a model can render it significantly more complicated than the original, and such scenarios should not be considered idealizations because they do not respect our pre-theoretical intuitions. See Hooker (1994) for a similar characterization of idealizations that also emphasizes the mathematical aspects. Historically, explicit and repeated use of idealizations as simplified distortions is traced back to Galileo (Clavelin 1974, Koertge 1977, McMullin 1985). That said, such notions arise already as far back as Plato's theory of forms, in which the concrete world is an approximate representation of an idealized world of forms.

Abstractions, on the other hand, have been most emphasized by Cartwright (1983; 1989, Ch. 5), and have been also called "negligibility assumptions" (Musgrave 1981), "method of isolation" (Mäki 1994), and "Aristotelian Idealizations" (Frigg & Hartman 2012). The idea here is that, rather than introducing falsehoods into our idealization, we instead strip away properties of the target system and ignore them (for similar characterizations, see also, for example, Brodbeck 1968, 460; Chakravartty 2001, 327; O'Neil 2000, 67-68). In other words, the abstracted analogue system is not, strictly speaking, a false description of the target system. Rather, it is partially true but not completely (or exhaustively) true.

Consider an example that illustrates the *interpretive* difference between the notions of idealizations and abstractions: We can think of a frictionless plane as an idealized system that stands in an idealization-relation to, and is approximately true of, a concrete plane with little friction, or is a false description of a plane with much friction. The frictionless plane is an idealization because we take it to be a distortion of the original concrete plane, which introduces a simplification. In contrast, we can instead think of a frictionless plane as an abstracted system, if we identify that friction is not essential to whatever phenomenon we are interested in and so we abstract it away in thought. The abstracted system stands in a relation of partial truth to the original concrete system. Historically, the method of abstraction (abstracting away the details) can be traced back to Aristotle and his philosophy of mathematics (Mendell 2004).

Approximations are meant to be a purely formal matter that is legitimized within the context of mathematics itself and the particular case at hand. Consider an example. It is a mathematical fact that a real-valued function $f(x)$ that is infinitely differentiable at a real number a can be expanded as an infinite power series, this being the Taylor series of the function:

$$f(x) = \sum_{n=0}^{\infty} \frac{f^n(a)}{n!}(x-a)^n = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots$$

(where $f(a)$ is the function's value at a, $f^n(a)$ is the n-th derivative with respect to x, $n!$ is the factorial of the integer n that is summed from zero to infinity, and the first three terms of the series have been written out.) Let us consider the particular case of the function $f(x) = e^x$ and expand it about the real number $a = 0$:

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + \frac{x^1}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

So far, our results have been exact. There have been no appeals to approximations. Now let us calculate the values of e^x when $x = 2$ for the exact case and compare it to various finite parts of the Taylor series. The exact value is $e^2 = 7.389056099 \ldots \ldots$, where values for the $n = 0, 1, 2, 3, 4, 5, 6$ cases are $1, 3, 5, 6.\overline{3}, 7, 7.2\overline{6}, 7.3\overline{5}$, respectively (for example, the calculation for the last term: $e^2 \approx 1 + \frac{2^1}{1!} + \frac{2^2}{2!} + \frac{2^3}{3!} + \frac{2^4}{4!} + \frac{2^5}{5!} + \frac{2^6}{6!} = 7.3\overline{5}$). The claim that e^2 is approximated by $7.3\overline{5}$ is grounded by the above procedure of expansion, while the further claim that this approximation is "good" or "good enough" will depend on the context of the case at hand. For example, if the observational accuracy of some empirically obtained measurement is such that there is ± 1 error associated with said measurement, then $6.\overline{3}, 7, 7.2\overline{6}, 7.3\overline{5}$ are all "good" approximations of 7.389056099 because they differ from it by ± 1. On the other hand, a calculation of asteroid trajectory might necessitate accuracy of several digits beyond the decimal point for which none of the above approximations are "good approximations." From a philosophical perspective, the main point is that, on the face of it, issues concerning physical interpretation need not arise. Accordingly, we tend to think that approximations do not pose any serious philosophical problems.

A noteworthy question (Ladyman 2008, 360-361), which has not received enough attention in the literature, is what distinguishes legitimate idealizations, abstractions and approximations from downright impossibilities? For example, idealized perfectly reversible Carnot engines, which could never be built in practice, are part and parcel of the theory of thermodynamics, while a perpetual motion machine (PPM) is absolutely prohibited by the same theory. There seems to be a distinction drawn here between a contingent possibility—possible but not practical—and a necessarily impossible one. However, it is not at all clear what grounds a distinction in which a Carnot engine is considered an idealization, while a PPM is considered an impossibility. One possible (conjectural) answer, appealing here to Norton's (2012) analysis of idealizations and approximations (see Chapter 5 and below), is that a Carnot engine arises as a structure that is asymptotically approached in some idealizing limit, while a PPM cannot be reached via such a limiting procedure.

This naturally connects with an additional distinction made recently by Norton (2012) between idealizations/abstractions and approximation.[62] The idea is that, whereas idealizations/abstractions bring with them novel reference—in particular, they refer to an idealized/abstracted analogue system—approximations are just "inexact descriptions" (Norton 2012, 209):

[62] Similar distinctions have been alluded to by Butterfield (2006, 24-5), Teller (1979, 348-9), McMullin (1985, 255).

An *approximation* is an inexact description of a target system. It is propositional.

An *idealization* is a real or fictitious system, distinct from the target system, some of whose properties provide an inexact description of some aspects of the target system.

It may be worthy to note that it is not clear to me how an object can be propositional, i.e., have semantic content, but yet not refer to anything (concrete or abstract). It seems to me that such an issue must be treated before the distinction can be embraced.

4.3 TAXONOMIES OF IDEALIZATIONS

The purpose of this section is to convey to the reader what are some of the most influential taxonomies of idealizations. It will be shown that the standard and most plausible justification for appealing to idealization is given by the sound principle, i.e. that in principle idealizations can be de-idealized away. We begin by discussing McMullin's (1985) classic paper on idealizations, and continue to present what can be thought of as a more modern treatment of McMullin's approach, a framework presented by Weisberg (2007a, 2013). Afterwards, we sketch another classical approach by Nowak (1980), and follow that up with a corresponding modern formulation by Shaffer (2012).

4.3.1 McMullin on Idealizations

One of the most influential taxonomies of idealizations appeared in Ernan McMullin's 1985 paper "Galilean Idealization." A Galilean Idealization is taken to signify the "deliberate simplifying of something complicated (a situation, a concept, etc.)," which may involve a distortion or may just leave aside some details (the former concept is what we have called an "idealization" and latter is we called "abstractions" in Section 4.2) (McMullin 1985, 248). According to McMullin (see Figure 4.1), "Mathematical idealization" imposes a mathematical formalism on a situation, and so the process of representing a concrete system in the outside world via mathematics already instantiates an idealization (McMullin 1985, 254). A "Galilean Idealization" comes in two flavors, the first, "Causal Idealization," is one in which the target of idealization is the "problem-situation itself," while the second, "Construct Idealization, is one in which the target of idealization is the "conceptual representation of the object." In other words, whereas, in the case of Causal Idealization, the target system is some concrete system, in the case of Construct Idealizations the idealization seems to inhabit two different levels: the target of idealization is a representation of a concrete system, that is to say, the target is some analogue system. The analogue system itself, arguably, is some sort of idealized representation of the original system, but we'll leave this complexity aside since McMullin does not attend to it.

Examples of Construct Idealizations might include simplified diagrams, as in geometry for instance, or mathematical equations. Construct Idealizations are created in two different manners and so they, in turn, also come in two forms, depending on whether idealized features are known to be *relevant*, in which we have a "Formal Idealization," or *irrelevant*, in which we have a "Material Idealization." So for example, a frictionless plane that models the behavior of a ball rolling down a plane is a Construct Idealization of the Formal Idealization kind, since friction is known to be relevant for the motion of the ball. However, if we further ignore the internal structure of the rolling ball or its color, because such features are irrelevant for the purposes of modeling the ball's motion, the Construct Idealization is of the Material Idealization kind. Clearly, a particular Construct Idealization can be both a Formal and Material Idealization (and also a Mathematical Idealization).

Causal Idealizations are ones in which the concrete system in the physical world itself is simplified. This can be done in two different forms. First, one could consider setting up an experimental procedure by which certain "idealizations" are instantiated. For example, we could conduct a ball-rolling-down-an-inclined plane experiment with different planes each with a decreasing amount of friction. This type of Causal Idealization is called "Experimental Idealization." One could also conduct such an experiment in thought and talk about "what would happen if…" such and such physical situations were instantiated. This second type of Causal Idealization is called a "Subjunctive Idealization." See Figure 4.1 for a summary of McMullin's taxonomy of idealizations.

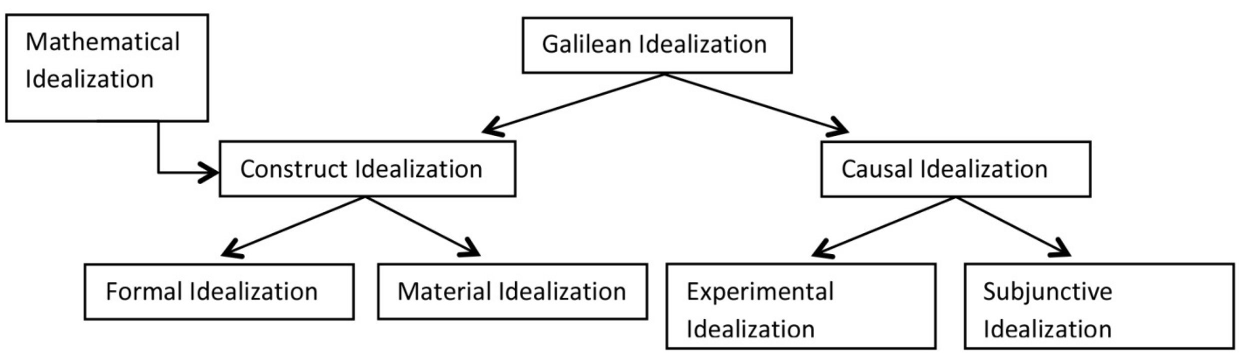


Figure 4.1: McMullin's taxonomy of idealizations.

How does one go about justifying the appeal to a Construct Idealization? According to McMulling (1985, 257), the justification includes the claim that, in principle, one can add back details to the model or representation without significantly damaging any sought after result: "…once the idealization has yielded a result, that result can (perhaps) be modified in order to make allowances for the 'departures from truth' that the original idealization required." Furthermore, this type of justification can be tracked back to Galileo (Drake 1967, 225):

> If we wish to use these conclusions proved by supposing an immense distance [from the earth's center] for finite distances, we must remove from the demonstration truth whatever is significant in [the fact that] our distance from the center is not really infinite.

The above process is standardly called "de-idealization" and the act of undertaking it "de-idealizing." A similar type of justification can be given for Causal Idealizations as is clear from the following well-known passage from Galileo (Drake 1967, 117):

> …let us (instead) observe what happens in the thinnest and least resistant media, comparing this with what happens in other less then and more resistant. If we find in fact that moveables of different weight differ less and less in speed as they are situated in more and more yielding media, and that finally, despite extreme difference of weight, their diversity of speed in the most tenuous medium of all (though not void) is found to be very small and almost unobservable, then it seems to me that we may believe, by a highly probable guess, that in the void all speeds would be entirely equal.

Important for our purposes, as we shall emphasize in Chapter 5, Galileo's justification for appealing to idealizations concerns looking to sequences of systems and inspecting how they behave in particular limits. It is exactly this type of justification that we will be investigating in concrete case studies in Chapters 6 and 7.[63]

4.3.2 Weisberg on Idealizations

Weisberg (2007a, 2013) takes an idealization to be the intentional introduction of distortion into scientific representation and presents a taxonomy of three kinds of idealizations: Galilean Idealizations, Minimalist Idealizations, and Multiple-model Idealizations (see Figure 4.2). He also identifies what he calls "representational ideals," which play a role in categorizing types of idealizations, and are made up of two components: "inclusion rules" and "fidelity rules." The representational ideals Weisberg identifies include "completeness," "simplicity," "1-causal," "maxout," and "p-general." *Completeness* inclusion rules dictate that all properties of target systems are to be included in its representation and completeness fidelity rules state that these properties be as similar as possible to that of the target system. *Simplicity* inclusion rules dictate to include as little as possible in a representation, while still being consistent with the fidelity rules, which demand a qualitative match between target system and its representation. *1-Causal* inclusion rules dictate to include only difference-making causal factors in a representation, and

[63] Shaffer (2012, 33) criticizes McMullin's taxonomy for two reasons. First, he claims McMullin does not seem to make room for the idealization/abstraction distinction. However, in defense of McMullin, one could claim that Formal Idealization incorporates the introduction of idealization through both removal and distortion of properties, as is clearly stated in McMullin (1985, 248, first full paragraph). In other words, although McMullin might not identify the distinction explicitly, *contra* Shaffer, there is room for it in the taxonomy. Second, Shaffer claims that Material Idealizations are not idealizations at all since "omitting *irrelevant* properties does not simplify a model in any useful sense" (Shaffer 2012, 33). The same criticism is put forward by (Liu 1999a, 246), who further criticizes McMullin, claiming that Formal Idealizations are types of Causal Idealizations since both seem to carve nature at it joints.

fidelity rules match those of simplicity. *Maxout* states that the precision and accuracy of a representation's prediction and retrodiction power should be maximized, without detailing how this should be accomplished. *Generality* as a representational ideal has two parts, *a-general* and *p-general*. A-general is the number of actual targets (systems or properties of systems), while p-general is the number of possible targets, that a particular model (or set of models) capture (Mathewson & Weisberg 2009). As an ideal, p-generality states that considerations of p-generality "should drive the construction and evaluation of theoretical models" (Weisberg 2013, 110).[64]

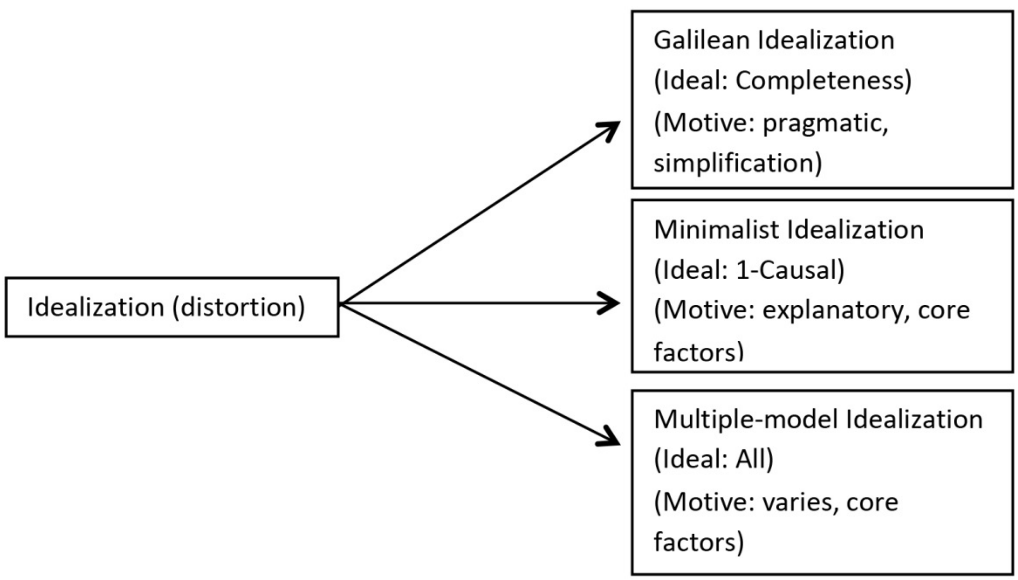


Figure 4.2. Weisberg's taxonomy of idealizations.

"Galilean idealization is the practice of introducing distortions into models with the goal of simplifying, in order to make them more mathematically or computationally tractable" (2013, 99). The motivation of tractability is a pragmatic one, and the justification given for Galilean Idealizations is that of de-idealization and concretization. The representational ideal of Galilean

[64] The entire presentation of representational ideals is a paraphrase of Weisberg (2013, 105-111).

Idealizations is completeness. This notion of idealization is attributed by Weisberg to McMullin (1985). However, Weisberg is clearly overly simplifying McMullin's account, which admits to a higher level of complexity. For example, McMullin's Causal Idealization, a type of Galilean Idealization, does not fit into Weisberg's characterization of Galilean Idealization, for Weisberg does not make a distinction between a distortion directed at a representation of a target system, and one directed at the target system itself. Moreover, it is questionable whether completeness as a representational ideal drives any scientific activity, let alone the construction of Galilean Idealizations. After all, if we abide by completeness, then it would seem that a perfect replica of a target would be the most complete representation of the target. However, why would a perfect replica of a target tell us anything about the target that the target didn't already tell us in the first place? Potentially, completeness is relevant to the construction of scale models, but even then there is a caveat. Specifically, we do not want the *scale* of the model to be driven by the representational ideal of completeness, i.e., the scale model and target must, at least, vary in scale.

"Minimalist idealization (also known as 'minimal model') is the practice of constructing and studying models that include only the core causal factors which give rise to a phenomenon," and are attributed by Weisberg to several different authors (Weisberg 2013, 100). For example, Strevens (2004, 2009) takes (what Weisberg calls) minimalist idealizations to be composed of difference-making causal factors, such that the removal of a causal factor prevents the entailment of a phenomenon of interest (for which the causal factors are difference makers). An "idealization" according to Stevens is the introduction of a non-difference making factor. Batterman (2002) and Batterman and Rice (2014) take minimal models to be those that are essential for the explanation of universal behavior such as critical phenomena, while Hartmann

(1998) seems to identify them with theories appealing to effective degrees of freedom. Weisberg also groups Cartwright's (1983; 1989, Ch. 5) account of abstraction under the category of minimalist idealization. The representational ideal for Minimalist Idealizations is 1-causal, although some worries concerning this claim will be worded below.

"Multiple-models idealization (hereafter MMI) is the practice of building multiple related but incomplete models, each of which makes distinct claims about the nature and causal structure giving rise to a phenomenon" (Weisberg 2013, 103). This type of idealization, with its corresponding modeling technique, is attributed mainly to Levins (1966), but also to sympathizers (Weisberg 2006, Matthewson & Weisberg 2009), and differs from Galilean Idealizations and Minimalist Idealizations "in not expecting a single best model to be generated" (Weisberg 2013, 103). While the motivation for Galilean Idealization is pragmatic and justified by de-idealization, the case is different, claims Weisberg, for minimal models and MMI. Minimal models are motivated by explanatory needs (e.g., Strevens' and Cartwright's causal explanatory models, Batterman's asymptotic explanatory model), while MMI are motivated by the varying goals of scientific theorizing, such as accuracy, precision, generality and simplicity. As for their justification, Weisberg explains, "…it is not justified by the possibility of de-idealization back to the full representation" (Weisberg 2013, 113). Due to the varying nature of MMI motivation, such idealizations can have any representational ideal.

Several comments are in order. First, there is a serious worry here that Weisberg's insistence on organizing various accounts according to his taxonomy significantly distorts the presentation of those accounts. For example, although both idealizations and abstractions might be appealed to with the goal of capturing core factors, the difference between introducing a distortion (idealization) and ignoring details (abstraction) might be philosophically relevant (see

Jones (2006) for such a claim). It is certainly historically relevant in the sense that abstraction refers to an Aristotelian technique (Mendell 2004 and references therein), while idealization refers to a Galilean one (McMullin 1985). Consider another example. Whereas Strevens identifies idealizations as non-difference-making causal factors, while minimal models are those that contain only the difference making causal factors, Batterman's discussion of minimal models takes the idealizations themselves to be essential to the explanatory story and, importantly, minimal models are non-causal on Batterman's account (Batterman and Rice 2014). This is worth emphasizing. There is no sense in which Batterman's account (of the appeal to minimal models for the purposes of the asymptotic explanation of universal phenomena) is a causal one, while Strevens' account is specifically causal.

Next at issue is the claim that with MMI there is no expectation of a single best model but with other types of idealizations there is. With respect to the latter part of the claim, there are clear counter examples. For instance, Batterman's asymptotic explanation of critical phenomena via Renormalization Group method appeals to classes of Hamiltonians and the manner by which they transform under the Renormalization Group transformation. In such a case, there is no reason to think that we are dealing with *one* idealized model. Rather, each Hamiltonian in a class of Hamiltonians refers to an idealized model with a possible corresponding idealized system. In what sense are all such Hamiltonians *one* minimal model instead of MMI? Moreover, in this case, the multiplicity of models does not stem from varying theoretical goals. Instead, there is *one* goal, of accounting for critical phenomena, in which all models play a role.

With respect to the first part of the claim, that with MMI there is no expectation of a single best model because varying theoretical merits necessitate a multiplicity of models, another objection can be raised. In particular, what is the argument for the claim that, in principle, no one model will be able to best account for all theoretical merits? Or that such a scenario is "unlikely?" Without such an argument, such in principle claims are unfounded. In particular, it does not follow from the fact that there are different theoretical merits, that indeed no one model best fulfills them all. Furthermore, any argument to the effect that there is no one best model is bound to set strict constraints on the future of scientific progress. History of science alone should convince us to reject any such constraints.

A last worry concerns the claim that minimal models and MMI, as opposed to Galilean Idealizations, cannot be de-idealized and "we should not expect it to abate with the progress of science" (Weisberg 2013, 103). There is something greatly worrisome about this claim. If it is truly the case that distortions and falsehoods are necessary to account for a phenomenon and, worse yet, portray the core causal factors entailing a phenomenon, how are we to interpret such an account realistically? The claim is not just that this would be a problem for a full-blooded scientific realist (see Section 4.4). Rather, anyone who does not want to commit oneself to an extreme instrumentalist view has to be able to tell a story about how, whatever work an idealization does for us, the ultimate success of a representation or model is that it gets something right about the world. In the case of causal minimal models, it seems such a story is at hand. However, in the case of the type of minimal models discussed by Batterman (2002), Batterman and Rice (2014), and Hartmann (1998), the story is more complicated than Weisberg's (2013) account admits (as I will stress in Chapters 5-7 and as has been recently emphasized by Batterman and Rice (2014)).

In other words, although Weisberg (2007a, 2013) rejects the sound principle, claiming that one need not appeal to de-idealization in order to justify appeals to MMI, I fail to see any good argument supporting this claim. It is true that science may have varying goals. However, it does not follow from the varying goals of science that there is some alternative justification for appealing to idealizations other than de-idealization as is discussed in the sound principle. At the very least, the burden of proof here clearly belongs to Weisberg (2007a, 2013). If not the sound principle, then what is the alternative justification for appealing to idealizations? I fail to find a plausible reply. Moreover, even if it is the case that many different models will be needed to account for different theoretical merits, each model justified independently will surely appeal to some sort of de-idealization scheme. So, for example, say I have two models, one that captures core causal factors and one used to make practical predictions. In each independent case we ought to expect a de-idealization process to justify appeals to idealizations.[65]

4.3.3 Poznan School on Idealizations

A different approach to idealizations in science is given by what has become to be known as the Poznań School methodology of science (Krajewski 1977, Nowak 1980; 1989).[66] In short, Nowak (1992) identifies five paradigms of idealizations in science. These include the neo-Duhemian paradigm, in which idealization is a method of transforming raw data, such as when systematic errors are corrected into data that can be used in the scientific enterprise (see also Suppes 1967). The neo-Weberian paradigm, in which idealization is a method of constructing scientific notions, and the neo-Leibnizian paradigm, in which idealization is a deliberate

---

[65] In the case of a model that captures core causal factors, our justification for idealizing away the non-core causal factors is that they are irrelevant to our subject of interest, and so, in principle, bringing such factors back into our model will not affect the result of identifying the core causal facts. In the case of a model that is used for making predictions, again, de-idealization ought to, in principle, only make the predictions more accurate. This is so because any idealization introduces error into our predictions. Unless one can give a de-idealization type story about how such error is not significant for our predictive purposes, the model's predictive success will remain a mystery.

[66] The literature on this approach is large enough to merit a chapter (or a dissertation) of its own, so I will not elaborate much on it here but instead I refer the reader to survey articles, Nowak (1992), along with authoritative compilations of collected papers: Brzezinski & Nowak (1992) and Nowakowa & Nowak (2000).

falsification. The neo-Millian paradigm, in which idealization is taken to be a consequence of the discrepancy between mathematical representation and the physical world. Last, the neo-Hegelian paradigm, in which idealization is the process of focus on only those essential and relevant features of some phenomenon.

In the Poznań School framework, one thinks of idealizations as "idealizational statements," which are conditional statements with an "idealizing condition" as the antecedent. For instance, a (material) conditional statement of the sort: "If the plane is a frictionless and infinitely long plane, then a body with initial velocity slides on it forever" is an idealizational statement with "frictionless and infinitely long plane" signifying the idealizing condition. What we called before "de-idealization" is called within this framework "concretization," and it is the process of removing idealizing conditions, replacing them with realistic ones, and accommodating for consequences that follow.

Consider a schematic exposition of the approach. Let T^0 be a law, model, or theory that entails (and governs) some phenomenon, and is a completely non-idealized representation of such phenomenon. Let T^k be a highly idealized law, model, or theory, with k idealized premises denoted by $p_i = 0$, which entails (and governs) some phenomenon, and only accurately represents those features that are essential, or relevant, for manifesting said phenomenon. Science is in the business of producing T^k, while philosophical issues (such as understanding scientific explanation and confirmation in light of ubiquitous appeals to idealizations) can be accounted for by the process of concretization via (conditional) idealizational statements in which we de-idealize T^k: $(T^k \& \, p_i \neq 0) \rightarrow T^{k-1}$, thereby creating a sequence leading to the non-idealized T^0: T^k, T^{k-1},…, T^0.

> A scientific law is basically a deformation of a phenomenon being rather a caricature of facts than generalization of them. The deformation of fact is, however, deliberately planned. The thing is to eliminate inessential components of it. (Nowakowa and Nowak 2000, 110)

For our purposes the main take-home message is that also on this influential approach, all idealizations and abstractions are ultimately justified view a de-idealization or concretization procedure, of the kind encapsulated in the sound principle.[67]

4.3.4 Shaffer on Idealizations

The view that the ubiquitous appeal to idealizations and abstractions in science renders scientific realism—roughly, the position that our best scientific theories are approximately true—untenable, is at the core of Nancy Cartwright's attack on scientific realism (Cartwright 1983, 1999) (see Section 4.4 below). Shaffer's (2012) work is a recent attempt to argue against such a line of thinking, and to make scientific realism compatible with the practice of idealization and abstraction in science. He suggests that there are two basic types of idealizations in science, *theoretical idealizations*, involving idealizations concerning theoretical claims, and *non-theoretical idealizations*, which concern idealizing initial and boundary conditions (Shaffer 2012, 14). Theoretical idealizations are further sub-divided into *non-constructive idealizations*, in

[67] An issue of controversy includes how to interpret the conditional in idealization statements. While Nowakowa and Nowak (2000) take idealization statements to be material conditionals, it has been argued by Niiniluoto (1990) (also see Shaffer 2001) that idealizational statements ought to be interpreted as counterfactual conditionals, e.g., "If it were the case that the plane was frictionless and infinitely long plane, then it would be the case that a body with initial velocity would slide on it forever." Nowak (Nowakowa and Nowak 2000) claims that such a move brings with it serious semantic and epistemological problems, but see Shaffer (2012, 69-81) for a reply.

which properties are idealized away by removal—call this *model-contraction* (what we have called "abstraction")—and *constructive idealizations*, in which properties are idealized away by replacing them with different properties—call this *model replacement* (what we have called "idealization") (Shaffer 2012, 33, 43). There are two types of non-constructive idealizations, *local non-constructive idealizations*, in which some parochial causal factor is idealized away as negligible, and *general non-constructive idealizations*, in which some causal factors are no longer parochial but are idealized away either because said factors introduce too much complexity—call these *strong general causal factors*—or because they are negligible compared to other primary causal factors—call these *weak general causal factors* (Shaffer 2012, 37, 39). (See Figure 4.3 for Shaffer's taxonomy of idealizations.)

All idealizations, however, are idealizational statements (see Section 4.3.3 above), understood as counterfactual conditionals, which are simplified representations of concrete target systems. Further, idealized representations are similar to their target with respect to structural, causal and/or dynamical features, for they allow surrogative reasoning about the actual world.[68] More precisely:

> A model M' is an idealization of a base model M if and only if M' is a simplified proxy for M such that M' represents M with respect to some of the features, $\{F_1, F_2, \ldots, F_n\}$, of M demmed to be scientifically interesting in some context C. (Shaffer 2012, 17, 91)

[68] See Swoyer (1991) and Chapter 3 for more on surrogative reasoning.

"Simplicity" (as well as "complexity") is understood in terms of properties and relations, and so it is fleshed out via Swoyer's (1991) notion of an "intentional relational system" (IRS).[69] The main idea is that it is possible to use IRS to compare two models M' and M, so as to decide if one is a simplified proxy of the other such that one might be an idealization (with respect to the other). Similarly, IRS can be used to cash out the representational relation between M' and M via "minimality," which states that some of the structure of the vehicle of representation M' that has empirical (i.e., measurable) content approximates some of the structure of the target of representation M that has empirical content (see Shaffer 2012, 92-3 for details). Moreover, models and truth-values of idealized models can be understood within the philosophical framework of possible worlds: A model or possible world is "real" if it is "complete," in the sense that every proposition in some language will be either true or false in that world. A model or possible world is idealized if it is an "incomplete world" or "partial model," in the sense that some proposition might not have a truth-value in such worlds, and so are taken to be false by default. Idealizations, then, are true simplified partial worlds that are most similar to the actual world. This is one of the main upshots of Shaffer's account since it allegedly allows him to reject the anti-realist claim—*à la* Cartwright (1983, 1989, 1999) (see Section 4.4 below)—which states

[69] Swoyer (1991) takes "structural representations" to be cases in which the vehicle and target of representation have a *shared structure* where, roughly, the idea is that the two be homomorphic to each other. The concept of shared structure is cashed out via what Swoyer (1991, 455) calls an *intensional relational system* (IRS)—an intensional kin of the logician's concept of a relational system. An IRS is an ordered 3-tupple $S_A = \langle I^A, \mathfrak{R}_n^{A,m}, f \rangle$ where I^A, $\mathfrak{R}_n^{A,m}$ and f are non-overlapping sets. I used the notation S_A because we think of an IRS as some (representation of) *system* A, with I^A its domain of individuals, $\mathfrak{R}_n^{A,m}$ is the domain of m^{th}-order n-place relations (i.e., $\mathfrak{R}_1^{A,1}$ is the set of first-order one-place relations or *properties*, $\mathfrak{R}_2^{A,1}$ is the set of first-order two-place *relations*, etc.) f is a unary function on $\mathfrak{R}_n^{A,m}$ that assign extensions, i.e., sets of individuals, to all the relations in this set (see Swoyer (1991, 501) for generalizations that include times and possible histories). We'll call the union of the domain of individuals and domain of relations the "total domain," and for simplicity we'll consider only first-order relations. Let h be a homomorphic function from the total domain of the IRS A to the total domain of the IRS B, which preserves the structure of relations in A, such that for every first-order n-place relation $R \in \mathfrak{R}_n^{A,1}$ and n-tuple of individuals $\langle i_1, \ldots, i_n \rangle \in I^A$: $\langle i_1, \ldots, i_n \rangle = f(R)$ if and only if $\langle h(i_1), \ldots, h(i_n) \rangle = f(h(R))$, where $S_B = \langle I^B, \mathfrak{R}_n^B, f \rangle$ is some other IRS B, $h(R) \in \mathfrak{R}_n^{B,1}$ and $\langle h(i_1), \ldots, h(i_n) \rangle \in I^B$.

that since theoretical claims in the sciences are only true in highly idealized models, they cannot be even approximately true of the concrete and actual world.

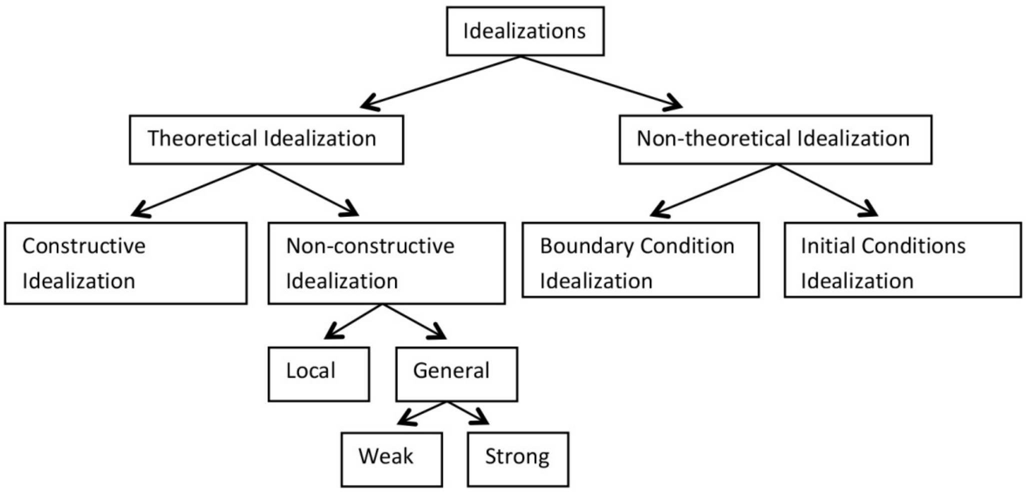


Figure 4.3: Shaffer's Taxonomy of Idealizations.

According to Shaffer (2012), idealizations are not only justified via de-idealization schemes. In fact, in his book he argues for the following thesis called the "weak ineliminability thesis" (WIT) (53):

> (WIT) Some idealizing assumptions . . . with respect to some theoretical claim T, are not even in-principle eliminable from T.

His main argument for the WIT is stated as follows (66-67; original emphasis):

> [It] is an empirical fact that the real – or actual – world is characterized by a variety of complex features and interactions. . . . It is also an empirical fact that human cognitive

> capacities and the computational resources of machines are considerably limited both individually and as a corporate body by both physical and computational constraints. So it is easy on this basis to argue that this disparity gives rise to a *physical* need on our part to impose simplifying assumptions on our descriptions of the world – these simplifications are known to be false – some of which we cannot in practice eliminate. On this view [WIT] can be grounded as a matter of physical necessity. It is simply a matter of empirical fact that most theories hold only under at least one idealizing assumption that cannot be eliminated . . .

In reply, and first, the claim that idealizations cannot be eliminated "in-principle" does not follow from the fact that they cannot be eliminated "in practice," but that seems to be the heart of Shaffer's argument above. Second, even if one can argue that there are senses for which idealizations are essential for science—as I will argue in Chapters 6-8—it does not follow from this claim that there exist *justifications* for appealing to idealizations that are significantly different from the sound principle and the de-idealization scheme inherent to it. We may have explanatory, methodological, and pedagogical reasons for appealing to idealizations, but still admit that a sound principle of interpreting physical theories would deem any effect that is an artifact of an idealization to be a non-physical or non-actual effect.

4.3.5 Conclusion: The Earman-Ruetsche Sound Principle

In conclusion, what I would like to emphasize is that, broadly speaking, all approaches to idealizations that we have encountered so far justify the appeal to idealizations via some de-idealization or concretization scheme, or else claim that an alternative justification is possible but then offer little to no convincing exposition of such an alternative justification. We may then place a principle of adequacy on an interpretation of the role of idealizations in science, namely, that such idealizations abide by the sound principle:

> EARMAN-RUETSCHE SOUND PRINCIPLE — If a scientific account (theory, model) uses an idealization to predict an effect which disappears when the idealization is removed then either the effect is an artifact of the idealization or else (if experiment confirms the effect) the theory is inadequate.

However, two worries arise. The sound principle seems to imply that idealizations are dispensable to science, e.g., that they are there solely for instrumental purposes, say, to make the math easier. This is worrisome because many believe that idealizations play substantive roles in science. Chapter 8 will be dedicated to alleviating this worry. There I will argue that the substantive roles of idealizations are compatible with the sound principle, so that there is a sense in which idealizations are essential for science, but still abide by the sound principle.

The second worry concerns the fact that identifying some particular structure in a theory as an "idealization," already makes substantial presuppositions about the ontology of the theory and the world. For example, many talk about a version of the thermodynamic limit, "the continuum limit" (Compagner 1989), as an idealization because such a limit is taken to represent a continuum, while objects in the physical world are fundamentally corpuscular, as the atomic theory of matter would suggest. However, such a claim makes some nontrivial assumptions about ontology, namely, that the ontology of our world is scale-independent, so that whatever things turn to be made up of, say, particles, waves, fields, etc., dictates the ontology of our world.

An alternative view takes ontology to be a scale-dependent issue. So, for example, we may say that on the macroscopic scale a drop of water *really* concerns a continuum, while on the microscopic scale it *really* is corpuscular, and there is no contradiction in the two statements. I am sympathetic to this view, but I will remain agnostic about it in this dissertation. Instead, what I want to emphasize is that *my embrace of the sound principle does not commit me to the view that ontology is scale-independent*. This issue will remain open.

4.4 IDEALIZATIONS IN PHILOSOPHY OF SCIENCE

The purpose of this section is to sketch how the pervasiveness of idealizations, abstractions and approximations in science raises various foundational problems in philosophy of science, specifically with respect to Laws of Nature, Scientific Confirmation, Scientific Explanation and the Realism/Anti-realism debate. It should be clear that the problems discussed become even more significant if idealizations appealed to are in some sense essential or indispensable.

The type of questions that we'll be dealing with consist of the following: If scientific accounts appeal to idealizations in accounting for concrete physical phenomena, i.e., they appeal to distorted analogue systems instead of the target systems being accounted for, is not the success of such accounts left mysterious? If two idealized hypotheses are incompatible with each other, which is to be taken as a guide to what the world is really like? If some effect depends essentially on an idealization that cannot be manifested in the laboratory, how is the effect to be empirically confirmed? If idealized law statements are falsehoods, how can they be confirmed by evidence? Moreover, if idealized laws are, strictly speaking, false, how is it said that they govern concrete behaviors of physical systems?

4.4.1 Laws of Nature

The last two questions have been raised explicitly by Nancy Cartwright in her well-known book, *How the Laws of Physics Lie*, which launched an attack on realist conceptions of laws of nature as fundamental and true.

> A long tradition distinguishes fundamental from phenomenological laws, and favors the fundamental. Fundamental laws are true in themselves; phenomenological laws hold only on account of more fundamental ones. This view embodies an extreme realism about fundamental laws of basic explanatory theories. Not only are they true (or would be if we had the right ones), but they are, in a sense, more true than the phenomenological laws that they explain. I urge just the reverse. (Cartwright 1983, 100)

Cartwright's goal here is to argue against the view known as *fundamentalism*—that there are true fundamental laws of nature and science that govern the behavior of matter throughout all of space and time—as well as to reject the received hierarchical connection between fundamental laws and the (non-fundamental) phenomenological ones. She calls the position that takes phenomenological laws to be true because they are logically entailed by the more general fundamental laws, the "generic-specific account" (Grünbaum 1954, 14; Cartwright 1983, 102). Roughly, Cartwright's claim is that the generic-specific account, and the fundamentalism position that drives it, cannot be true because in order to derive phenomenological laws from fundamental ones, one has to appeal to approximation schemes and idealizations that are, strictly speaking, false, and so such derivations cannot be sound.

Moreover, bracketing aside worries about the generic-specific account for the time being, fundamentalism itself is objectionable on grounds that fundamental laws are only true, strictly speaking, in highly idealized or abstracted scenarios that never obtain in concrete physical situations and so, *contra* the fundamentalist, all fundamental laws are really only *ceteris paribus* laws:

> My conclusion from looking at a large number of cases of how theories in physics are used to treat real situations in the world, both in testing the theories and in their impressive technological applications, is that it is always *ceteris paribus* regularities that come into play. All the cases I have looked at have just the characteristic I point to: they are either especially engineered or especially chosen to include only those causes that occur in the preferred set of the theory. They are, moreover, always arranged in a very special way: a way that the theory knows how to describe and to predict from. That is not surprising where *ceteris paribus* laws are involved, since we can neither test laws of this kind nor apply them until we are sure the *ceteris paribus* conditions are satisfied. The point is that these are the kinds of cases that give us our most powerful reasons for accepting our theories in physics. And the laws they give us reason to accept are all *ceteris paribus* laws. (Cartwright 2000, 210)

Similar sentiments are conveyed in Cartwright (1999, Ch. 2). For our purposes we do not need to dive into the *ceteris paribus* debate (see Earman *et al.* 2002). Rather, what is important to identify is that all of Cartwright's talk (in the above quote) of "especially engineered," "especially chosen," "arranged in a very special way" and "all things being equal" amounts to an appeal to idealized conditions, abstracted scenarios and approximation schemes. Thus, if laws given to us by actual scientific practice "are all *ceteris paribus* laws" then by the same token, they are all laws that appeal to idealizations, abstractions and approximations. In response to such concerns some authors have suggested that fundamental laws do not govern concrete systems in the world, but instead govern abstract or fictional models or scientific representations (e.g., Cartwright 1983, 1999; Giere 1999a, 1999b; van Fraassen 1989). Still other authors such as Teller (2001) agree that idealizations are an indispensable part of science, thereby seemingly defending Cartwright's thesis:

> Indeed, simplifying is just what physics and most other sciences do… In 1974 I read through all of Feynman's Lectures on Physics (1963). I was flabbergasted. Almost nowhere could I find deductions from first principles. Most of the work involved ingenious tailoring of both the laws and facts to produce accounts which, at the same time, fit the world well enough but were also sufficiently simple for us to manage. (Teller 2001, 394)

However, Teller does not come to the same conclusion as Cartwright for he takes truth itself to be qualified, inexact or partial (Teller 2001, 2004, 2011). This amounts to rejecting standard accounts of truth as exact correspondence.

In the following section we will turn to the issue of scientific confirmation and idealizations, but before doing so let me reiterate the take home message of this section: The appeal to idealizations in deriving phenomenological laws from fundamental laws, as well as the alleged necessity of referring to idealized conditions in order for fundamental laws to be, strictly speaking, true, threaten fundamentalism (or a realist conception of laws).

4.4.2 Scientific Confirmation

A different problem that arises by appeals to idealizations concerns scientific confirmation and this can be the confirmation of theories, laws, scientific hypothesis, models, etc. The problem is nicely demonstrated within the context of Hempel-Oppenheim's hypothetico-deductive (HD) method (Hempel and Oppenheim (1948), reprinted in Hempel (1965)), although it can be generalized to any scheme that goes from a theory that is to be tested, on the one hand, to the experimentally testable results, on the other hand, by way of auxiliary assumptions. In short, the idea is that idealizations and approximations act as types of auxiliary assumptions so that all the problems that auxiliary assumptions raise for confirmation theory are also raised by idealization and approximation. The difference is that, whereas auxiliary assumptions tend to be unnoticed and become prominent when there is predictive failure, idealizations and approximations are known to be false right from the start. The point is illustrated by Laymon (1985, 1989) and I shall follow him closely in this subsection.

Consider some scientific account S that is composed of a set of premises T, such as a theory and/or law(s) and/or boundary condition(s), and a set of various idealizations, abstractions and approximations I, which are known to be, strictly speaking, *false* premises. Let P be some directly or indirectly observably verifiable prediction or retrodiction, and assume that P follows from T and I (either in a deductive manner or by some mathematical derivation) so that $T \,\&\, I \rightarrow P$. There are two options: either experiment shows us that P or that not P. If the former, i.e., the scientific account predicts empirically adequate results, then we know T must be *false* because only a false, say, theory, when conjoined with false idealizations could lead to correct predictions. If the latter, then we know that either both T and I are false, or T is true but I is false, or T is false but I is true. However, since I is known to be false, the use of idealizations, abstractions and approximations effectively shields T from being disconfirmed. That is to say, we know I is false but we don't know whether T is true or false. We don't know if our failure to confirm the theory arises because of problems with the theory T itself or because of our appeal to idealizations I. In order to remedy the situation, one modifies the confirmatory scheme to take the notion of approximate truth into account:

CONFIRMATION		DISCONFIRMATION	
$T \& I \rightarrow P$	*is true*	$T \& I \rightarrow P$	*is true*
I	*is approximately true to degree* d	I	*is approximately true to degree* d
P	*is approximately true to degree* $S(d)$	P	*is **not** approximately true to degree* $S(d)$
T	*is true*	T	*is false*

In other words, a theory is confirmed when P is predicted by a scientific account S to some degree $S(d)$ computed on the basis of S and it is disconfirmed when P is not true to degree $S(d)$. However, herein lies "the crux of the problem:"

> Idealizations are introduced precisely because we lack the analytic and computational skills to calculate "correct" theories, but in the absence of these correct theories we cannot compute the predictive bias introduced by our idealization… the assumption that we can meaningfully assign some degree or measure of the distance of our idealizations

[or approximate predictions] from the truth [, independent of a correct scientific account,] … is not generally true. (Laymon 1985, 153-4)

And so it is not clear how to make sense of scientific confirmation when appeals to idealizations, approximations, and abstractions are made by candidate scientific accounts. Said differently, we generally believe that if a scientific account S is correct, then we are justified in believing conclusion P that follows from the account. But how can we have justified belief in P when it follows from a set of premises T & I that include falsehoods?[70]

One might worry though that the above problem posed to confirmation theory by idealizations is somewhat outdated and does not extend to approaches that prevail today, such as Bayesian confirmation theory (Bayesianism). This is not the case. Shaffer (2001) (see also Shaffer 2012, Ch.3) challenges (subjective) Bayesianism to show that idealized hypotheses have some degree of confirmation, thereby pressing for a proposal for how to assign posterior probabilities to counterfactual conditionals.[71] If h is a hypothesis and e the evidence, $\mathrm{P}(h \mid e)$ is the posterior probability of h given e, and plays a role in determining how well e confirms h. There have been various suggestions made by Bayesian theorists as how to flesh out "e confirms h"[72] but all such suggestions involve the quantity $\mathrm{P}(h \mid e)$. If idealized hypotheses are to be construed as counterfactuals as Shaffer insists, wherein $>$ is the symbol for counterfactual

[70] Liu (2007) criticizes Laymon's possible solution to the idealization problem *a la* de-idealization. See Shaffer (2012, Ch. 4) for an account of scientific confirmation via inference to best explanation, which accommodates involved idealizations. Davey (2011) appeals to a type of contextualization to explain of how justified beliefs can be generated from appeals to idealizations, abstractions and approximations.

[71] See Jones (2006, Ch. 7) for a Bayesian reply to Shaffer (2001).

[72] Some options of what "e confirms h" amounts to include, among others, a difference measure such as $P(h \mid e) - P(h)$ (Earman 1992 and Rosenkrantz 1994), a log-ratio measure such as $\log(\frac{P(h \mid e)}{P(h)})$ (Howson and Urbach 1993, and Milne 1996), a counterfactual difference measure such as $P(h \mid e) - P(h \mid \neg e)$ (Christensen 1999 and Joyce 1999); see Hartmann and Sprenger (2010) for a Bayesian primer.

conditional, A is an idealized antecedent condition and C is a consequence, then $\mathrm{P}(h \mid e)$ becomes $\mathrm{P}(A > C \mid e)$. According to Bayes' theorem,

$$\mathrm{P}(A > C|e) = \frac{P(e \mid A > C)P(A > C)}{P(e)}$$

But it is well-known that there exists no nontrivial and coherent proposal for assigning values to $P(A > C)$ (Shaffer 2001, 43-45). Hence, Shaffer concludes (2001, 45, his emphasis),

> [The] situation is unfortunate for the Bayesian as there does not seem to be any extant, coherent, suggestion as to how we are to nontrivially assign prior probabilities to indicative or counterfactual conditionals. This problem, *the Bayesian problem of idealization*, appears to have devastating consequences for Bayesianism. Unless Bayesians can come up with a coherent suggestion for how such probabilities are to be understood, either Bayesianism must be rejected or, given the ubiquity of idealizations, Bayesians must accept the rather counterintuitive conclusion that few, if any, scientific theories have ever been confirmed to any extent whatsoever.

4.4.3 Scientific Explanation

Idealizations also raise problems for accounts of scientific explanation. For instance, our discussion of the problems posed by idealizations to the HD method of confirmation, which is an extension of Hempel and Oppenheim's (1948) deductive-nomological (DN) account of explanation (also known as the "covering law" account), can be extended to any type of nomological account of explanation wherein the appeal to idealized laws is made use of in the explanation of phenomena. This is so because, according to DN the explanans that explains phenomena must be true, but idealized explanantia are false (Hempel 1965). In similar manner, causal approaches to explanations such as Lewis' (1986), Salmon's (1998) and Woodward's (2003), and unification accounts such a Kitcher's (1981), all require factual correctness and deny falsehoods as part of their explanatory approach. See Jones (2006, Ch. 1) for an elaboration on how idealizations pose problems for traditional representative accounts of scientific explanation, such as ones just mentioned, which fail to incorporate falsehood into their explanation schemes.

However, there are accounts that do attempt to incorporate idealizations into their approach to scientific explanation and these usually go under the name of "Explanatory Falsehoods" or "Explanatory Fictions." What is particularly interesting is that most accounts of explanation, insofar as they can be modified to accommodate the inclusion of idealizations, depend crucially on one's ability to de-idealize said idealization. Four such accounts include (Jones 2006; 35, 58):

1. Ronald Laymon [(1980), which] allows idealized descriptions to be explanatory if they counterfactually approximate correct descriptions.
2. Alexander Rueger and David Sharp [(1998), which] allow idealized descriptions to be explanatory if they qualitatively approximate correct descriptions … [in the sense that] (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.[73]

[73] See Rueger and Sharp (1998, 212) for more on "structural stability."

3. Philip Kitcher [(1989), which] allows idealized descriptions to be explanatory if the error due to the idealized descriptions is either negligible or unlikely to make a non-negligible difference.
4. Ronald Giere [(1988)], R.I.G. Hughes [(1993)], and Paul Teller [(2001), which] allow idealized descriptions to be explanatory if the systems they describe are sufficiently similar to real systems.

All of these explanatory accounts seem to accommodate the appeal to de-idealizable idealizations, i.e., idealizations that are consistent with the sound principle. That said, see Jones (2006) for a defense of the claim that none of the above accounts can accommodate "ineliminable idealizations," i.e., what we have been calling "essential idealizations."

4.4.4 Scientific Realism

"Scientific Realism," roughly, can be taken as the position that science "aims to give us, in its theories, a literally true story of what the world is like; and acceptance of a scientific theory involves the belief that it is true (van Fraaseen 1980, 8).[74] Wherein, "to have good reason for holding a theory is *ipso facto* to have good reason for holding that the entities postulated by the theory exist" (Sellars 1963, 97). Moreover:

---

[74] The situation is in fact more complicated since there are various dimensions to scientific realism. Roughly, the metaphysical dimension has to do with commitment to the existence of a mind-independent world and/or objects within it, (possibly) along with their properties and relations. The semantic dimension concerns interpreting the claims of scientific theories literally as satisfying truth conditions. The epistemic dimension regards the empirical and explanatory success of science as evidence for the (approximate) truth of the claims of science, so that scientific claims constitute knowledge about the world and objects within it. The axiological dimension takes the aim of science to give approximately true descriptions and faithful representations of the world. See Boyd (1983, 45), Psillos (1999, xix), Niiniluoto (1999, 21) and Chakravartty (2011).

> Theoretical terms in scientific theories (i.e., non-observational terms) should be thought of as putatively referring expressions; that is, scientific theories should be interpreted "realistically." … The reality which scientific theories describe is largely independent of our thoughts or theoretical commitments. (Boyd, 1984, 41)

However, appeals to idealizations and abstractions seem to undermine scientific realism as has been stressed by Cartwright (1983, 1989, 1999) and formulated in deductive form by Hughes (1990, 71):[75]

> Premise 1. Scientific theories provide models of the processes and entities of nature.
>
> Premise 2. To have a model, whether of a physical process like the propagation of light or of an entity like an atom, is not to have a literally true account of the process of entity in question.
>
> Conclusion. It follows that science does not provide true descriptive accounts of these processes and entities.
>
> Corollary. If we consider theory alone, scientific realism cannot be justified.

Furthermore, consider the two main (and related) arguments in favor of scientific realism include the "Ultimate (or Miracle) Argument" and an appeal to "Inference to Best Explanation" (IBE). The Ultimate Argument for scientific realism states that realism "is the only philosophy that doesn't make the success of science a miracle."

---

[75] See Shaffer (2012) and Pincock (2012) for a reply.

> That terms in mature scientific theories typically refer …, that theories accepted in a mature science are typically approximately true, that the same term can refer to the same thing even when it occurs in different theories—these statements are viewed by the scientific realist not as necessary truths but as part of the only scientific explanation of the success of science, and hence as part of any adequate scientific description of science and its relations to its objects. (Putnam 1975, 73)

However, if scientific theories, laws, and models appeal to idealizations, abstraction and approximations, and this includes appealing to fictional entities, how is the Ultimate argument cogent? That is to say, *if the success of science can only be explained by believing in the existence of the entities science posits, and if science ubiquitously appeals to fictional entities known not to exist* (through idealizations, abstractions, approximations) *in achieving its "success," then the ultimate argument is reduced to absurdity for it compels us to maintain that various entities are and are not fictions.* A similar problem arises when IBE is taken as an argument for realism.

> Let us suppose that we have evidence E, and are considering several hypotheses, say H and H'. [IBE] says that we should infer H rather than H' exactly if H is a better explanation of E than H' is… It is argued that if we follow this rule in all 'ordinary' cases; and that if we follow it consistently everywhere, we shall be led to scientific realism, in the way Sellar's dictum [(above)] suggests. (van Fraaseen 1980, 19)

Yet, three problems arise when one considers idealizations. First, as discussed in the last subsection, how can an explanation be explanatory if some of its explanatia are idealizations and thus false? Second, granting that we can make sense of idealized explanation, how are we to tell which parts of a scientific explanation is descriptively true and which parts are false? That is to say, how do we know *beforehand* which explanatia are idealizations and which tell us what the word is really like? Third, given that both *H* and *H'* appeal to idealizations so that neither is strictly speaking true, how can we tell which one is a better guide to ontology?[76]

In contrast to the aforementioned problems posed by idealizations to realism, some authors consider the practice of idealization to support a realist conception of science. For example, commenting on the process of de-idealization, McMullin (1985, 262, 264) notes that

> …the 'adding back' [, i.e. de-idealizing,] if it accounts for additional experimental data and especially if it leads to the discovery of new empirical laws, is a strong validation for the model and its accompanying theory… this becomes a strong (though not conclusive) argument for the existence of structure postulated by the model… What makes it heuristically sensible to proceed in this way is the belief that the original model *does* give a relatively good fit to the real structure of the explanandum object. Without such a fit, there would be no reason for the model to exhibit this sort of fertility. This gives perhaps the strongest grounds for the thesis of scientific realism.

Similar sentiments can be found in Nowak (1980) and in Brzezinske & Nowak (1992), and in the many works of Ronal Laymon (1980, 1984, 1985, 1989a, 1989b, 1995). Laymon (1985, 155)

[76] See Sklar (2000, 2003) and Chapter 3.

begins by noting that we can account for the process of scientific confirmation via our ability to de-idealize with the “converging counterfactual theory of confirmation:”

> A scientific theory is confirmed (or receives confirmation) if it can be shown that using more realistic idealizations will lead to more accurate predictions.
>
> A scientific theory is disconfirmed if it can be shown that using more realistic idealizations will not lead to more accurate predictions.

He then argues that cases in which de-idealization leads to “successful convergence to better experimental fit are *miraculous coincidences* for the antirealist” (Laymon 1984, 118). Antirealists, e.g., Cartwright (1989), argue that it might not be possible to improve a model by de-idealization, and that the procedure is not faithful to actual scientific practice, in which scientist tend to shift to different models (Hartmann 1998). Furthermore, it is claimed that sometimes multiple contradictory models are used to represent the same phenomenon (Morrison 2000).

In conclusion, the ubiquitous appeals to idealizations, abstractions, and approximations in science raise various philosophical problems. Taking a coarse-grained perspective, we may say that all such problems can be alleviated by appealing to standard de-idealization or concretization schemes, as is encapsulated in the sound principle.[77] If, however, there were idealizations that were inconsistent with the sound principle, then all the problems that arise due

[77] From a more fine-grained perspective, even if idealizations are consistent with the sound principle, certain technicalities need to be accounted for with respect to particular theories of scientific laws, explanation, confirmation, and specific accounts of scientific realism.

to the presence of idealization would reemerge. In the following chapter we turn to a generic discussion of such idealizations, and in Chapters 6-7 we look at purported concrete examples.

5.0 PATHOLOGICAL IDEALIZATIONS

The goal of this chapter is to make use of distinctions, due to Butterfield (2011) and Norton (2012), between *limit properties* and properties of *limit systems* in order to characterize the notion of a *pathological idealization*, specifically, an essential idealization that is inconsistent with and marks the failure of the sound principle.

5.1 LIMIT PROPERTIES AND LIMIT SYSTEMS

The sound principle tells us that if we remove an idealization—call this process *de-idealization*—then those effects that disappear cannot be counted as genuinely physical effects. We can gain further insight into the meaning of the principle if we relate the notions of introducing and removing idealizations to limiting procedures, where taking a limit may correspond to introducing an idealization or abstraction. In a recent work, John Norton (2012, Section 3) distinguishes between a "limit system" and a "limit property."[78] A limit system concerns a system and its limit, i.e., a limit of a sequence of systems, while a limit property has to do with the property of a system and its limit, i.e., a limit of a property of a sequence. The

[78] We could equally well talk about "objects" and "relations." A system is an object. System properties are one-place relations. Parts of the system, or parts of the object, are other objects that stand in part-whole relations to the system, and so on.

distinction is best understood through some simple examples, which we shall turn to soon. First though, let us add some precision to our discussion.

Specifically, Butterfield (2011, Section 3) discusses similar distinctions. He makes a distinction between a *system* $\sigma(N)$ that depends on some parameter N (let $\{\sigma(N)\}$ denote a sequence of such systems), a *quantity* defined on the system $f(\sigma(N))$ (let $\{f(\sigma(N))\}$ denote a sequence of quantities on successive systems), and a (real number) *value* $v(f\big(\sigma(N)\big))$ of quantities on successive systems (where a sequence of states on $\sigma(N)$ is implicitly understood; let $\{v(f\big(\sigma(N)\big))\}$ denote a sequence of values on successive systems). A *limit system* $\sigma(\infty)$ arises when $\lim_{N\to\infty}\{\sigma(N)\}$ is well-defined—otherwise there is no limit system. A *property of a limit system* refers to the value $v(f\big(\sigma(\infty)\big)$ of the (natural) limit quantity $f\big(\sigma(\infty)\big)$ (in the natural limit state) on $\sigma(\infty)$. A *limit property* $v(f\big(\sigma(N)\big)$ is a limit of a sequence of values of quantities on successive systems (or, values on the systems on the way to the limit) and is well-defined when $\lim_{N\to\infty}\{v(f\big(\sigma(N)\big))\}$ exists. The question that I will be discussing is whether a property of a limit system equals the system's limit property. More precisely, the question asks whether $v(f(\sigma(\infty)) = \lim_{N\to\infty}\{v(f\big(\sigma(N)\big))\}$ (assuming both are well-defined).

Let us now illustrate the distinction with an example, which will also foreshadow problems to come. Envision a concrete three-dimensional cubical system with sides of length L and with some impenetrable point-like object in its interior (i.e., a "hole"). We wish to represent this system, so we call it our *target system.* The target system is the concrete and physical system in the world, which we wish to study. Pictorially, one may imagine a scientist in a box conducting experiments (see Figure 5.1). The object that represents the target system will be called the *vehicle* of representation. A corresponding "realistic" or "faithful" representational vehicle would consist of modeling the target system as a three-dimensional cubical space with an impermeable hole (see Figure 5.2a).[79]

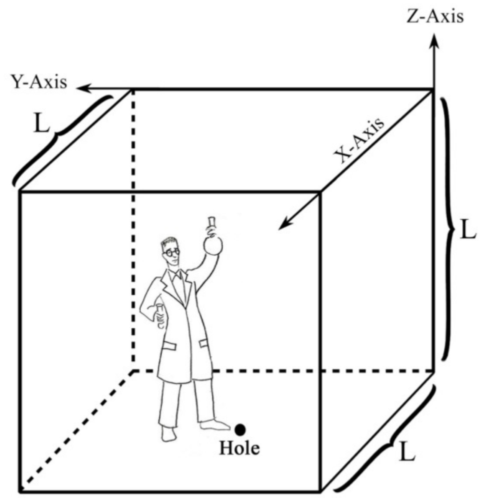


Figure: 5.1 The target system. A scientist in a box, with an impenetrable object represented by a "hole," conducting experiments.

We'll be interested in two properties concerning this target system (or its faithful representation). First, whether or not loops inside the cubical space can be continuously deformed (shrunk) into a point. If this is possible we'll say that (the topology of) the space is *simply connected*, that it has a *trivial topology*, and that its *fundamental group is trivial* (see Figure 5.2b). If not, we'll say the space is *non-simply connected* (or that it is *multiply connected*), that it has a *non-trivial topology*, and that its *fundamental group is non-trivial* (see Figure 5.3a).[80] Second, we'll want to ascertain whether the cubical system has the course-grained

[79] Roughly, by a "faithful representation" I mean a vehicle that can be used to extract inferences about a target that are true of the target (Contessa 2007), but also that the vehicle can tell us what the target is like (in the sense of Sklar's (2003) "guides to ontology"). I discuss the notion in detail in Chapter 3.
[80] These are all rough and intuitive characterizations of the notions of connectedness, homotopy, etc, and will do for my purposes. For precise characterization see standard textbooks on topology and algebraic topology, e.g., Munkres (2007), Hatcher (2002).

property of a non-zero or exactly zero (three-dimensional notion of a) volume V. Clearly, our cubical system does have a non-zero volume equal to L^3 and (as can be seen in the Figure 5.3b) its space is simply connected since all loops−even those that lie in the plane with the impenetrable hole−can be continuously deformed and shrunk to a point.[81]

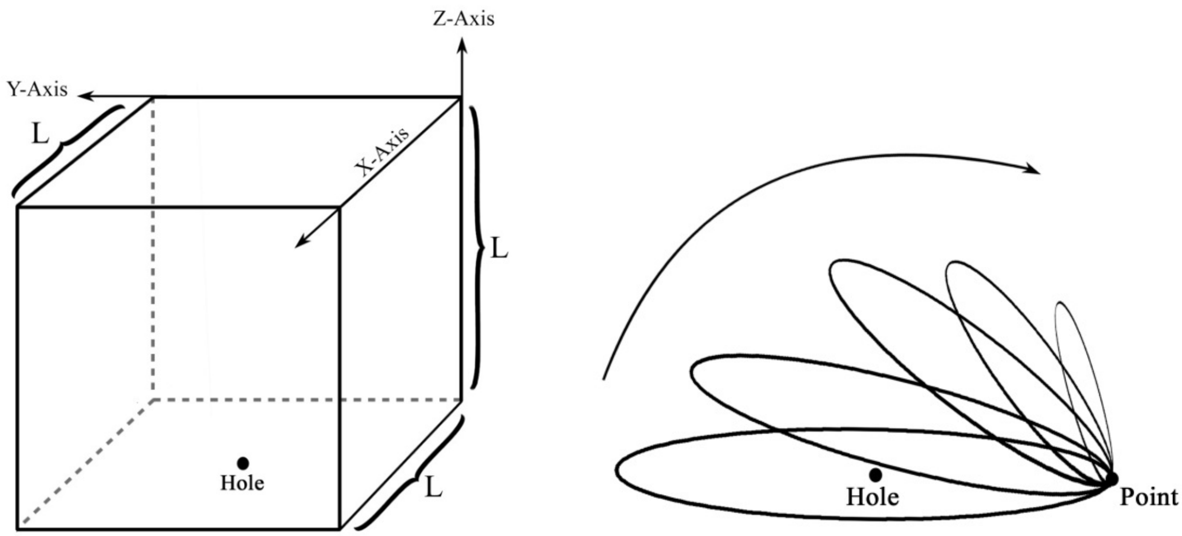


Figure 5.2: (Left) (a) A 3D cubical space with a hole (an impenetrable obstruction). (Right) (b) A simply connected space, with a trivial fundamental group, in which loops can be continuously deformed to a point.

Next, consider a family of such systems in which the height of the cuboids, denoted by l, shrinks (see Figure 5.4). That is to say, we'll consider the sequence in the limit in which $l \to 0$, and the location of the interior hole in the sequence will be a function of the height of the cuboid. So long as we are *on the way to* the limit system, each system in our sequence will have some non-zero volume and will be simply connected. However, *at the limit* our limit system will be a

---

[81] The volume will clearly be equal to L^3 because the Lebesgue integral is invariant under the replacement of a function with one that is almost identical in every way, i.e., one that differs only on a measure zero subset.

two-dimensional square with a zero volume and a multiply connected space (see Figure 5.5). That is to say, loops surrounding the hole cannot be continuously deformed to a point without being obstructed by the impenetrable hole (see Figure 5.6). We have here a case in which the limit properties, i.e., a simply connected space with a trivial fundamental group and a non-zero volume, do not match the corresponding properties of the limit system, i.e., a multiply connected space with a non-trivial fundamental group and a zero volume. A lack of correspondence between limit properties and properties of limit systems marks a potential failure of the sound principle, or, said differently, shows that the limit system fails the Earman-Ruetsche soundness test. In other words, when we partially remove the idealizations by de-idealizing away from the limit system we lose the properties of zero volume and a non-trivial topology. Now, insofar as such properties are used to predict a physical effect that has been empirically confirmed, the removal of the idealizations renders the effect non-physical according to the sound principle, and so a problem arises: an effect that is non-physical according to the sound principle is actually empirically confirmed in the laboratory—and so the sound principle fails.

Concentrating on the multiply connected topology of the space, I call such an idealization a *pathological idealization*. It is an *idealization* because we are dealing with limiting procedures, particularly, with a property of a limit system.[82] It is a *pathological* idealization because the property of the limit system does not correspond to the limit property (in the sense that any de-idealization from the limit systems renders the property nonexistent), and because there is no sense in which the multiply connected space can be seen to emerge continuously in the limiting procedure. Said differently, *the idealization is pathological because there is no sense in which a*

[82] One may wonder why my example of a sequence of cuboid-models ought to be labeled an "idealization," when I have not identified what in the physical world is being idealized. My point is that given some physical cubical system with an impenetrable object in its interior, all the cuboids in the sequence except for the first are idealized models of this original physical system. See Butterfield (2011) and Norton (2012), for further details and discussion.

non-simply connected topology is the limit of simply connected topologies. To see this, contrast the topology of the space with the property of zero versus non-zero volume. Such a coarse-grained distinction would have us think that that the zero volume of the limit system is also the product of a pathological idealization. But if we instead look at the more fine-grained *value* of the volume, it is clear that a property of the sequence is that the volume goes to zero.[83] Said differently, if we de-idealized away a small epsilon ε amount from the limit system so that the height $l = \varepsilon$, then the volume will be $V = L^2\varepsilon$ and this is approximately equal to zero.[84] In contrast, the space will not be "approximately multiply connected" on the way to the limit system, and we cannot create a space that is as "approximately multiply connected" as we may want by choosing an appropriate epsilon ε. Instead, the space will be simply connected, full stop. See Table 5.1 for a summary.

Limit Property:	Property of Limit System:	Type of Idealization:
Simply Connected Space	Multiply Connected Space	Pathological
Volume $V = L^2\varepsilon$	Zero Volume $V = 0$	Non-Pathological $V = L^2\varepsilon \approx 0$

Table 5.1: Summary (of the example) exemplifying the distinction between limit property and limit system.

[83] What I mean by transitioning from a coarse-grained to a fine-grained distinction, in Butterfield's (2011) terminology, is to transition from looking at a particular value of a quantity on successive systems to a different one that sheds light on the first. See Butterfield (2011, 1078-1079) for an illuminating and simple example.

[84] By "approximately" we mean that we can have the volume $V = L^2\varepsilon$ be as small as we want by choosing an appropriate epsilon. Said differently, for any non-zero epsilon of maximal error about the volume one may care about, we can find a de-idealization with volume less than this epsilon.

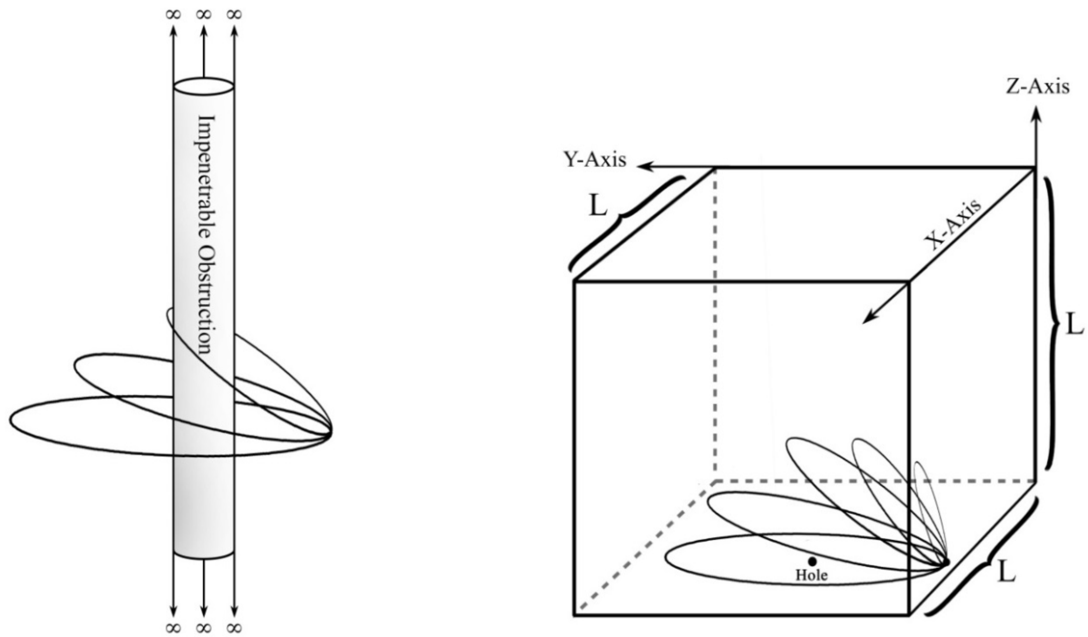


Figure 5.3: (a) (Left) Multiply connected space, with a non-trivial fundamental group, in which some loops cannot be continuously shrunk to a point. (b) (Right) A cubical space that is simply connected and has a volume of $V = L^3$.

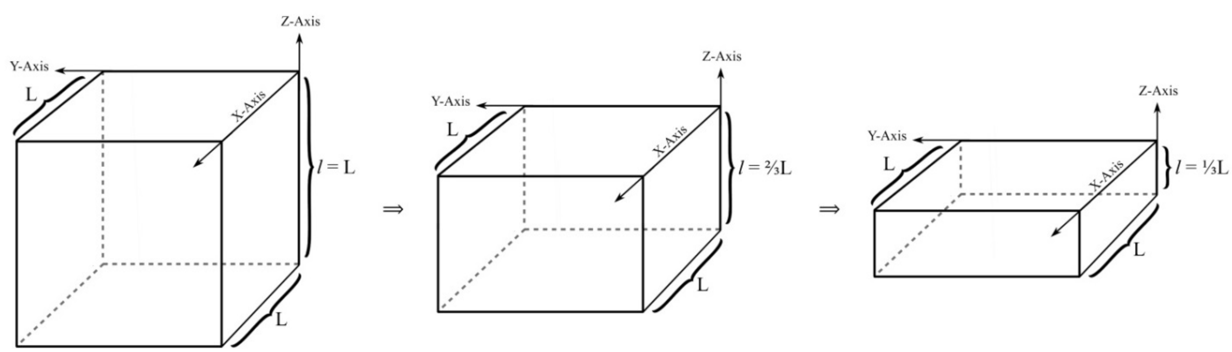


Figure 5.4: A sequence of systems of cuboids as the height shrinks.

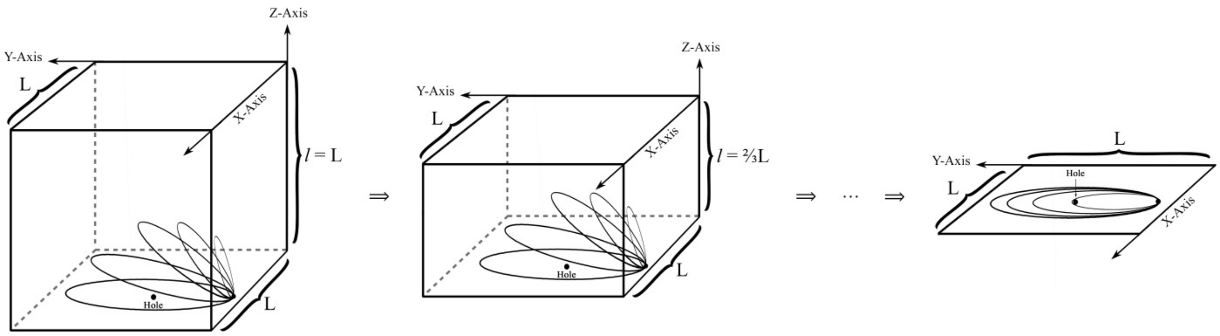


Figure 5.5: On the way to the limit, each system in the sequence of cuboids is simply connected with a trivial fundamental group and a non-zero volume. At the limit, the system is multiply connected with a non-trivial fundamental group and a zero volume.

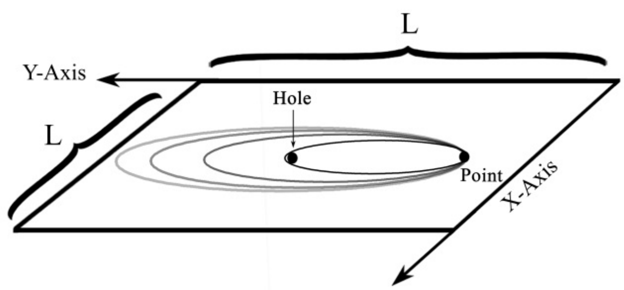


Figure 5.6: At the limit, the limit system is a two-dimensional square and so the topology of the space is multiply connected, with a corresponding non-trivial fundamental group, because loops surrounding the hole cannot be shrunk to a point.

5.2 THREE TYPES OF INCOMPATIBILITIES

Norton (2012) identifies three different situations that might arise when we compare limit properties and limit systems:

1. Limit system and limit properties are compatible. (212-213)
2. There exists no well-defined limit system. (213)
3. Limit system and limit properties are incompatible. (213-214)

In the context of the first case we have a justification for appealing to an idealization via the sound principle since when we de-idealize the property we are interested in does not disappear. The second is in some sense less interesting since there is no well-defined limit system, no idealized analogue system, which may cause problems for the sound principle. The third case marks the potential failure of the sound principle and it is this case that I would like to elaborate on.

Specifically, let us consider *three types of incompatibilities* between limit properties and the corresponding properties of limits systems.

3i. Incompatibility to some degree delta δ.

3ii. Incompatibility due to a coarse-grained distinction.

3iii. Strict incompatibility.

All three cases of incompatibilities concern potentially problematic idealizations, but only case 3iii is a pathological idealization that marks the failure of the sound principle. This is so since, for 3ii, we can alter our perspective in more of a fine-grained manner so as to demote the incompatibility to compatibility (as in the case of zero and non-zero volume discussed above).

In the case of 3i, in which the incompatibility exists to some degree delta δ—say, the limit property p differs from the corresponding property of the limit system p' such that $p = p' + \delta$—we can justify the appeal to the idealization instrumentally.[85] That is to say, we allow ourselves to make use of the idealization, admitting that it is not a faithful representation of the states of affairs, but it is "good enough" for the context and purpose we are interested in.[86]

However, in the context of 3iii, there is no clear sense in which the limit property and the property of the limit system approximate one another, and it is not the case that one can bring about compatibility by looking to a different property (as in the zero versus non-zero volume case). This is further emphasized with examples in Chapter 6 (Aharonov-Bohm effect) and Chapter 7 (fractional quantum statistics). Thus, while idealizations, generally, are compatible with the sound principle, pathological ones are not. Accordingly, they ought to either be dispensed with, or else, future theory will show that what we thought was an idealization, in fact, is not (in which case, again, we have salvaged the sound principle).

It is worthwhile to place a particular caveat on this last statement. The sound principle concerns the prediction and explanation of concrete effects in the world that appear in some target system (or systems) of study. It is in this context that we must dispense with pathological idealizations. However, and first, it does not follow from this claim that non-pathological idealizations and abstractions cannot play an indispensable role for explanatory purposes. Second, nor does it follow that one cannot appeal to pathological idealizations for methodological and pedagogical purposes, wherein there is no appeal to a concrete target

[85] In short, it is a combination of the techniques described in 3i and 3ii that alleviate any mystery associated with the conventional definition of phase transitions.

[86] Norton (2012, 213-214) considers such an example. Here the property we are interested in is the ratio of the surface area to volume of a sphere elongated as an ellipsoid. While the limit property is $\frac{3\pi}{4} \approx 2.35619449$, the corresponding property of the limit system is 2. Thus, if an error ± 0.4 of is admissible in some particular context both the limit property and the (incompatible) property of the limit system will be "good enough."

system.[87] As I hinted at in Chapter 3 (in the context of deciphering a representational code) and will argue for in Chapter 8, even with the dispensation of pathological idealizations in an explanatory context, there are substantive methodological and pedagogical roles for pathological idealizations to play in science. Moreover, I will show how the well-received taxonomies of idealizations discussed in the last chapter distort, rather than shed light, on the story behind the effects that I will be looking at and the idealization appealed to in such a context.

5.3 CONCLUSION

To summarize, the ubiquitous appeal to idealization, abstractions and approximations in the context of scientific accounts of physical phenomenon raises a host of philosophical problems. Most such problems can be solved—or great headway can be made in solving said problems—by noting that such idealizations can be, in principle, de-idealized so that they are consistent with the sound principle. This is the standard and most plausible justification for appealing to idealizations in science. Essential idealizations are interesting because they seem, on the face of it, to be inconsistent with the standard justification. However, this chapter shows that only pathological idealizations are problematic. Reflecting on the case study of first-order phase transitions discussed in Chapters 2 and 3, we may now identify that phase transitions concern non-pathological essential idealizations. That is to say, although, strictly speaking, phase transitions are defined only for systems with well-defined thermodynamic limits, we can think of

[87] In McMullin's (1985) terminology, as long as we are dealing with a construct idealization, which does not represent any concrete target systems in the world, but only an abstract system (existing in our mind or in a platonic heaven), then there is no inconsistency with the sound principle since any effect that will be predicted will concern the abstract system, namely, the construct idealization itself.

these *theoretical* constructs as *abstract* phase transitions that offer unfaithful (but sound) representations of concrete phase transitions that occur in the world. Appeals to the thermodynamic limit are justified instrumentally via 3i and 3ii (above). That is not to say that the thermodynamic limit plays a purely instrumental role in the discovery, study and characterization of phase transitions. Chapter 8 will identify its substantive value (albeit, in a different context).

This ends our general discussion of scientific representation (Chapters 2 and 3) and idealizations (Chapter 4 and this chapter). I now move onto several concrete case studies and flesh out some implications.

6.0 IDEALIZATIONS IN THE AHARONOV-BOHM EFFECT

This chapters looks at the nature of idealizations and representational structures appealed to in the context of the Aharonov-Bohm (AB) effect. It is suggested that interpreting the effect as fundamentally topological in nature commits one to an untenable view of the necessity of idealizations in science. An alternative account is outlined and endorsed. (Implications for the debate revolving around essential idealizations, asymptotic explanation, and minimal models, are discussed in Chapter 8.)

6.1 INTRODUCTION

In Chapters 4 and 5, we discussed the nature of idealizations in science and identified pathological idealizations, i.e., idealizations that are inconsistent with the sound principle, as ones that ought to be dispensed with. In this chapter I will show that the received view of the AB effect appeals to pathological idealizations and thus does not conform to the sound principle.

The structure of this chapter and my argument is as follows. Section 6.2 will sketch the received view of the AB effect and explain in what sense the nature of idealizations appealed to are topological. In short, the received view holds that in order to account for the AB effect one must appeal to an idealized multiply connected configuration space (or a base space on the fiber bundle formulation). I explain in Section 6.3 that this is problematic for it deems the AB effect a consequence of a pathological idealization that does not conform to the standard justification via the sound principle. I thus reject the standard story in Section 6.4 by presenting a non-topological interpretation of the effect and dissolution of the problematic pathological idealizations. I also consider an objection to my account based on the fiber bundle formulation of the effect (for it is in this context that such objections have been raised in the literature). I end the chapter with a short summary in Section 6.5. Technical appendices relating to the AB effect can be found at the end of the dissertation.

6.2 THE RECEIVED VIEW OF THE AB EFFECT[88]

When a beam of charged particles is shot from a source at a double-slit screen and then made to recombine at some detector-screen, an interference pattern emerges akin to the type one would expect to see from waves interfering with each other (see Figure 6.1). Standard (non-relativistic) quantum mechanics (QM) describes such states of affairs with a *wave function* Ψ that represents the (quantum) state of the system. It turns out that in the vicinity of a magnetic field produced by a solenoid the interference pattern is shifted (see Figure 6.2). Mathematically, this is represented by (the two components of) Ψ incurring an appropriate (non-trivial, relative) *phase factor* $e^{i\theta}$ under its natural temporal evolution governed by Schrödinger's equation—the main dynamical law of QM.[89] *What accounts for this phenomenon?*

[88] My discussion of the AB effect brackets many technical details to a set of appendices at the end of the dissertation. Appendix A discusses canonical quantization for a charged particle in electromagnetic fields, and Appendix B presents the standard textbook account of the AB effect (such as it arises in, e.g., Ballentine (1998, 321-325), Griffiths (2005, 368-390)). Appendix C sketches the fiber bundle formulation of the effect.

[89] A bit more precisely, we may follow Ruetsche (2011, 20-21) and identify the majority of QM theories as ones in which particular states of affairs are described in terms of a system's state and observable physical magnitudes

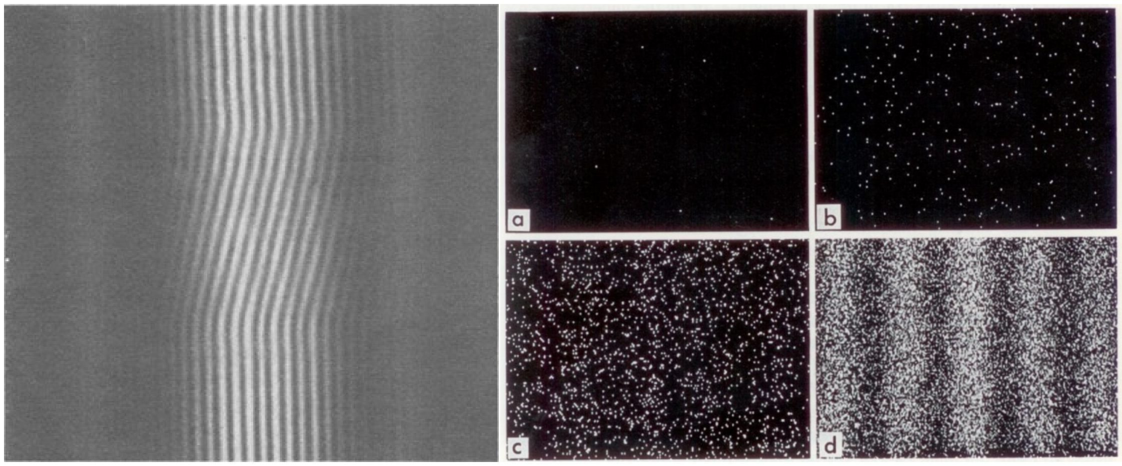


Figure 6.1: (Left) An example for an interference pattern from a double-slit experiment (from Möllenstedt and Bayh 1962, 304). (Right) Single-electron build-up of (biprism) interference pattern (from Tonomura 1999, 15). (a) 8 electrons, (b) 270 electrons, (c) 2000 electrons, and (d) 60,000 electrons.

In a famous article, Aharonov and Bohm (1959) derived the inference shift by making use of a particular *Hamiltonian* that I will denote with H_{AB}^{I}.[90] A Hamiltonian corresponds to the energy observable (i.e., a physical magnitude) pertaining to a system and is of special importance for it generates the temporal evolution of the system's state. H_{AB}^{I} contains what is known as the

(where states and observables correspond to the system's kinematics), as well as the dynamics governing temporal evolution. QM represents observable physical magnitudes with self-adjoint elements of a set of bounded operators acting on a separable Hilbert space $\mathcal{H}$ (i.e., $\mathcal{H}$ has a countable basis). States are represented by a set of density operators (positive trace-class operators of trace 1) on $\mathcal{H}$ (i.e., states—understood as normed, positive, countably additive assignments of probabilities to projection operators—stand in a one-to-one correspondence with density operators). It then turns out that, given some plausible assumptions, which I will not elaborate on here, a wave function is a special case of a more generic quantum state that corresponds to a "pure" state (as opposed to a "mixed" state). According to the "Schrödinger picture," the dynamics of a system in an initial state $W(0)$ (at time $t = 0$) are governed by $W(t) = e^{-iHt}W(0)e^{-iHt}$, where H is the Hamiltonian operator (roughly, the energy observable) pertaining to the system. $\mathcal{H}$ is chosen as the space of square-integrable complex-valued functions. On the "Heisenberg picture," it is the observables that evolve with time (and the state remains static): $A(t) = e^{-iHt}A(0)e^{-iHt}$, where A is an observable. Here $\mathcal{H}$ is chosen as the space of square-summable infinite sequence of complex numbers. The term e^{-iHt} is the time-evolution operator, where the Schrödinger equation is the special case of the dynamical law governing a wave function (a pure state).

90 Aharonov and Bohm (1959, 486) used the standard Hamiltonian for a charged quantum particle in classical electromagnetic fields, which is appropriate for use in the region outside the solenoid (see Appendices A and B). I use the notation H_{AB}^{I} in order to emphasize that this is the *idealized* AB effect Hamiltonian.

electromagnetic vector potential $\boldsymbol{A}$ (or *vector potential* for short), which is a mathematical entity from which the magnetic field $\boldsymbol{B}$ can be derived, and is intimately connected to the phase factor $e^{i\theta}$. Said differently, H_{AB}^{I} is the representational structure that is *fundamental* for it allows one to derive the state of the system Ψ, as well as whether or not it will incur a (relative) phase factor $e^{i\theta}$ leading to a shifted interference pattern.

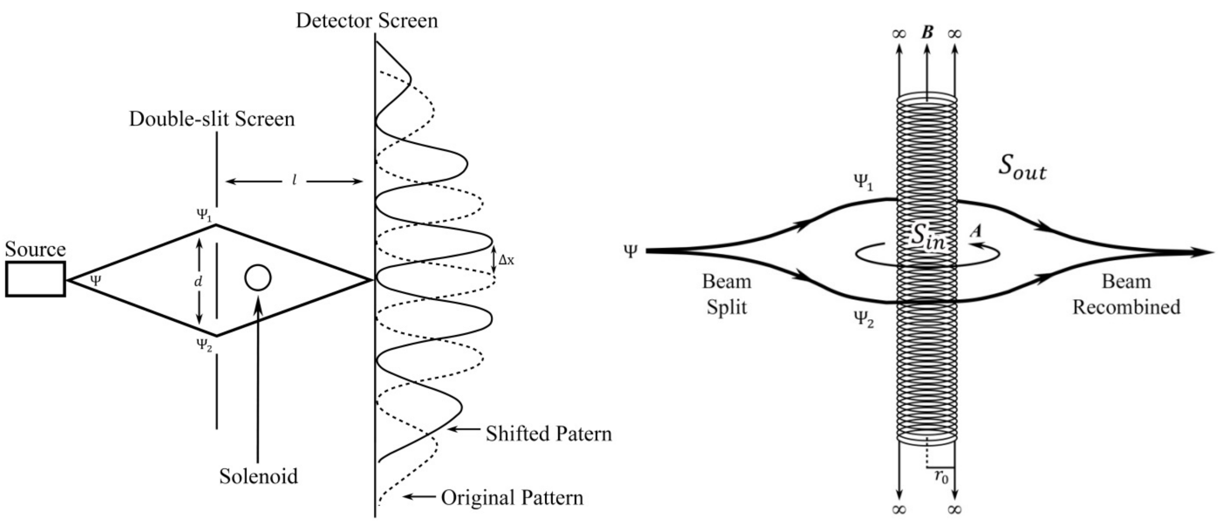


Figure 6.2: The AB effect. A beam of electrons Ψ is split in a region S_{out}, made to encircle a solenoid (that generates a magnetic field inside the region S_{in}), and then to recombine on a detector screen. The original interference pattern is shifted by an amount Δx. See Appendix B.

The effect became known as the "Aharonov-Bohm effect."[91] It is of special interest to philosophers because it seems to portray a type of quantum non-locality (since the magnetic field

91 To be exact, this is the magnetic AB effect. There is an analogue electric AB effect, implied by Lorentz covariance, which arises from electric (instead of magnetic) fields. See Peshkin and Tonomura (1989) for a review of the effect and its experimental confirmation, along with a selective review of the debate that arose with respect to the reality of the effect in the physics literature. Also see Tonomura (1999, Ch. 6; 2010) for a recent overview of

affects the beam while being confined to a region where the beam is not) and raises a host of interpretive issues regarding ontology and indeterminism in classical electromagnetism and quantum mechanics.[92]

Our interest will center on the idealizations that arise in the received account of the effect. We want to understand what grounds the particular choice of Hamiltonian used by Aharonov and Bohm (1959), beyond the fact that it allowed them to predict the AB effect, which was subsequently empirically confirmed.[93] Roughly, H_{AB}^{I} is taken as the appropriate Hamiltonian with which to represent the system under heavily *idealized* conditions. The solenoid giving rise to the magnetic field is assumed to be infinitely long so that the magnetic field remains within the solenoid and does not leak out. It is also assumed that the solenoid is absolutely impenetrable so that the wave function does not penetrate the region containing the magnetic field. Together, such idealized assumptions can be incorporated into one: *The topology of the space in which the AB effect manifests is multiply connected*.[94] More precisely, the assumption of absolute

---

experimental confirmations and application of the effect. Ehrenberg and Siday (1949) are usually credited with first noting the effect and Chambers (1960) with being the first experimental confirmation. Tonomura et al. (1982, 1986) is considered the first definitive confirmation of the effect. Also see Caprez *et al*. (2007) for recent experimental confirmation emphasizing the absence of a force in the manifestation of the effect (i.e., there is no unknown force, such as the Lorentz force, shifting the electron wave packet).

[92] See, for example, Batterman (2003), Belot (1998), Healey (2007), Leeds (1999), Lyre (2001, 2004), Maudlin (1998), Mattingly (2006), and Nounou (2003) for some of the philosophical literature. Most of this literature concerns the physical reality or unreality of the vector potential. I should emphasize that this topic, for which there is no "received view," is *not* the subject of this chapter.

[93] As Tonomura (1999, Ch. 6) explains, the definitive confirmation of the AB effect by Tonomura *et al*. (1982, 1986) did not make use of infinite and impenetrable solenoids. Instead, superconducting toroidal solenoids were used. Such solenoids are prepared so as to minimize magnetic field leakage and maximally shield the solenoid from an electron beam with a copper and niobium coating. A shift in interference pattern was observed. Although part of the shift could be explained in terms of the unideal conditions and other sources of error, these aspects could not account for the entire shift in the pattern. For example, it was calculated that the magnetic field leakage can affect the phase of an incident electron beam by at most π, but the observed relative phase shift was about $12\,\pi$. For these reasons there is little doubt now that the AB effect is a real physical effect and not an artifact of idealizations. That being said, the conceptual foundations of the effect are not well understood in the sense that it isn't clear what *causes* the effect. The debate in the philosophical literature (see footnote 92) revolves mostly around this last issue.

[94] Speaking more carefully, we would say that the QM formulation of the AB effect concerns a Hilbert space of continuous functions on a non-simply connected configuration space constituted by $\mathbb{R}^3$ with an infinite cylinder

impenetrability of the infinite solenoid, or absolute impenetrability for toroidal solenoids of the kind used in the experimental confirmation of the effect (by Tonomura *et al*. 1982; 1986), renders *the configuration space available to the wave function representing the beam of electrons non-simply connected*. For example, in the original Aharonov and Bohm (1959) paper, the "Discussion of Significance of Results" begins as follows:

> The essential result of the previous discussion is that in quantum theory, an electron (for example) can be influenced by the [vector] potentials even if all the field regions are excluded from it. In other words, in a field-free *multiply-connected region* of space, the physical properties of the system still depend on the [vector] potentials. (490; emphasis mine)

Similarly, Peshkin and Tonomura (1989, 27) note that in order to obtain the conditions for a manifestation of the electric AB effect, one "requires a multiply-connected spatial geometry, so there can be no electric AB effect in a simply-connected region." Ryder (1996, 102) goes so far as claiming that it is "an essential condition for the Bohm-Aharonov effect to occur that the configuration space of the vacuum is not simply connected."

Relatedly, ever since Wu and Yang's (1975) influential paper on the AB effect, it has become common to appeal to fiber bundle formulation of classical electromagnetism to discuss the effect (see Table 6.1).[95] In this context, electromagnetic fields are represented by the *curvature* of, and the electromagnetic vector potential is represented by a *connection* on, the

removed: $\mathbb{R}^3 \setminus S_{in}$ (where $\mathbb{R}$ is the set of real numbers). For my purposes this changes little since the problem and "source of ambiguity in the theory is the multiple connectedness of the domain" (Magni and Valz-Gris 1995, 179).

[95] See Healey (2007, Ch. 1-2, Appendix B) for more thorough (but philosophically friendly) introduction. Also see Appendix C for an overview of the essentials.

principal fiber bundle appropriate for the formulation of classical electromagnetism.[96] The relative phase factor $e^{i\theta}$ that gives rise to interference pattern arises as the (non-trivial) *holonomy* (or *anholonomy*) of a closed curve C encircling solenoid.[97] Details aside, it is standardly claimed that in order to get a non-trivial holonomy, the fiber bundle *base space* that represents physical space (or spacetime) must be *multiply connected*. The reason is that vanishing electromagnetic fields around the solenoid correspond to a curvature that is zero. Zero curvature means that "the connection on this bundle is flat everywhere in this region" (Healey 2007, 42). Moreover, if "there is a nontrivial holonomy [as is the case with the AB effect] . . . and if the connection is flat, *the base space* [representing physical space] *must be nonsimply connected*" (Batterman 2003, 542; original emphasis). Thus, the "ultimate 'cause' of the [interference] shift is the topology of the base manifold . . ." (Nounou 2003, 193).

We then may summarize the received view of the nature of idealizations that arise in the context of the AB effect, along with the standard explanation, as follows:

> THE RECEIVED VIEW: In order to account for the AB effect it is necessary to idealize the space in which the effect manifests as topologically multiply connected. (This usually takes the form of a *non-simply connected electron configuration space* due to the dual idealizations of an infinite and impenetrable solenoid appealed to by the original Aharonov and Bohm (1959) formulation,[98] or a multiply connected base space on Wu

---

[96] That is to say, a principal bundle, where the base space is the spacetime manifold, and where the structure group is $U(1)$—the multiplicative group of complex numbers of modulus 1 (i.e., the group of rotations in the complex plane).

[97] That is to say, observable effects are produced by elements of the holonomy group determined by the connection on the principal fiber bundle.

[98] In the context of toridial solenoids, all that is needed is the impenetrability assumption.

and Yang's (1975) fiber bundle formulation.[99]) It is in virtue of the non-trivial topology that there exists an AB effect.

	Electromagnetic Vector Potential	Magnetic Field Produced by Solenoid	Shift In Interference Pattern (due to a Non-Trivial Phase Factor)	Space or Spacetime
Non-Relativistic Quantum Mechanic Formulation	$\boldsymbol{A}$	$\boldsymbol{B}$	$e^{i\theta}$ $= \exp(\frac{iq}{\hbar}\oint_C \boldsymbol{A}\cdot d\boldsymbol{r})$	$\mathbb{R}^3$ Or $\mathbb{R}^4$
Fiber Bundle Formulation	Connection	Curvature	Non-trivial Holonomy	Base Space

Table 6.1: Comparison of terminology between non-relativistic quantum mechanics and the fiber bundle formulation of the AB effect.

Sentiments of this sort arise in, among others, Aharonov and Bohm (1959), Batterman (2003), Belot (1998, 544), (Healy 2007, 42), Lyre (2001, 2004), Morandi (1992, V), Morandi and Menossi (1984), Nakahara (1990, 356-359), Nash and Sen (1983, 301), Nounou (2003), Peshkin and Tonomura (1989), Ryder (1996), Schulman (1971), and Wu and Yang (1975; 3845).

[99] I do not think this is the position that Wu and Yang's (1975) proposed. Rather, it seems to be the one advertised by various philosophers (e.g., Batterman 2003, Nounou 2003, Healey 2007).

6.3 THE PARADOX OF THE AB EFFECT

The received view is untenable. To get a sense for why this is so, consider the following set of propositions that the standard approach aspires to embrace:[100]

1. Real systems are simply connected.[101]
2. Real systems display the AB effect.
3. The AB effect occurs if and only if there is a non-trivial phase factor.
4. A non-trivial phase factor arises if and only if a system is multiply connected.[102]

While the first two propositions imply that real systems are simply connected and display the AB effect, the last three propositions convey that real systems are multiply connected in virtue of displaying the AB effect. We then have a paradox: real system are and are not multiply connected (or, real systems do and do not display the AB effect). In other words, something has gone horribly wrong. Thus, one can justifiably question the legitimacy of the idealizations appealed to in the context of the AB effect:

> It is a general attitude in physics that any mathematical model of reality should react "continuously" to small changes in the formulation of the problem. Accepting this

---

[100] I am drawing here an analogy with the case study of phase transitions discussed in Chapter 2. Proposition 1 is discussed in the example of a system of cuboids discussed in Chapter 5, and proposition 4 is the one that I ultimately reject.

[101] Said differently, the configuration space available to the electron beam is the entirety of $\mathbb{R}^3$ (where $\mathbb{R}$ is the set of real numbers), such that loops can be continuously shrunk to a point. Proponents of the received view standardly admit that appeals to multiple connectedness concerns an idealization, thereby committing to proposition 1.

[102] More precisely, by "multiply connected" we mean a non-simply connected configuration space (or a non-simply connected base space on the fiber bundle formulation).

> principle, as rigid a topological property as multiple connectedness should be watched carefully… After all, infinitely repulsive barriers do not really exist and all that is needed is a theory that can describe the experimental facts when the repulsive barrier is sufficiently high. (Magni and Valz-Gris 1995, 179-180)

Said differently, on the received view of the AB effect the idealizations appealed to in order to account for the effect are topological in the sense that one needs to appeal to a multiply connected space. However, we saw in Chapter 5 that such a topological idealization is pathological, for a multiply connected topology is a property of a limit system that does not match the corresponding limit property—any minute de-idealization renders the topology of the space simply connected. This is also the case for systems manifesting the AB effect: as long as the solenoid is finite and penetrable, the electron configuration space corresponding to the AB effect will be simply connected (see Figure 6.3). It is only at the limit that the property of multiple connectedness arises. Connectedness is a binary topological property so it makes no sense to talk about "approximate multiply connectedness," or to state that the physical space "approximately instantiates" the structure of a non-trivial topology. Either the space is connected or it is not; there is no intermediate. In the same vein, multiple connectedness does not arise continuously in some limiting procedure. Limiting procedures already presuppose topological notions such as continuity.

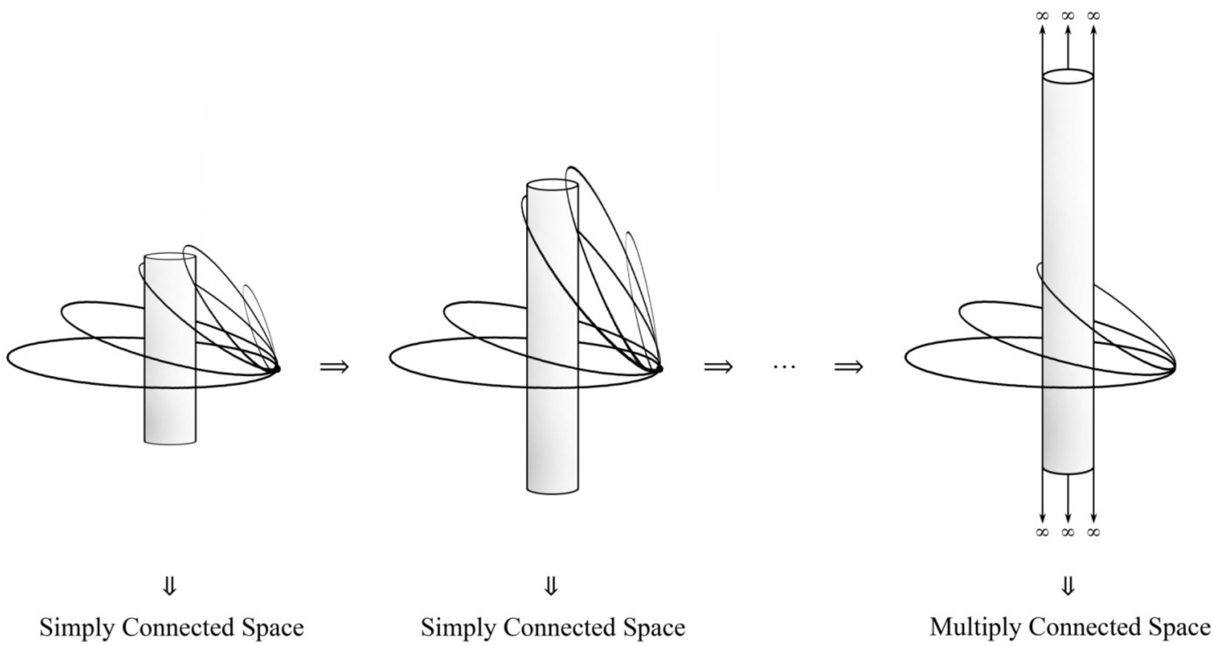


Figure 6.3: A schematic illustration of why the received account of the AB effect appeals to topological idealizations.

Nevertheless, an objection arises. Is it not possible that some piece of fancy mathematics, some new technique or other, will allow for a precise formulation of an "approximately multiply connected" space so that my qualms about the received view will be unfounded? Maybe there is a way to put topologies on sets of geometries such that a multiply connected space *is* the limit of a sequence of simply connected spaces? I believe that such an objection may be evaded in two steps. First, note the position that I am defending, which explicates why such techniques are not available, is the typical one in the sense that a precise characterization of "approximately multiply connected" is not part and parcel of the literature on topology. Thus, I submit that the burden of proof (to offer such a characterization that accommodates an account of the AB effect interference shift) is not my own, and belongs to a defender of the received view. Second, if such techniques were available, then they would only support the same position that I am defending here. Specifically, I claim that topological idealizations that are *pathological* ought to be avoided. If new techniques avail manners by which to account for the AB effect, by appealing to non-pathological topological idealizations, then so much the better. Such techniques can be subsumed under the alternative view that I espouse in the following subsection. What is certain is that such new techniques are *not* part of the received view.

In sum, if we side with the current received view we commit ourselves to an untenable position regarding the necessity of idealizations since we are in effect saying that certain properties of a limit system—that do not correspond to anything in reality—somehow account for the AB effect. I urge against taking such a route because a more sober approach is available.

6.4 JUSTIFYING THE AB EFFECT: A NON-TOPOLOGICAL INTERPRETATION

On the received view, the AB effect is accounted for by the Hamiltonian H_{AB}^{I}, which is the result of highly idealized conditions (namely, infinite length and absolute impenetrability), entailing an appeal to a multiply connected electron configuration space. The philosophical question of interest is whether one can justify the appeal to H_{AB}^{I} without appealing to a non-trivial topology. That is to say, we want to avoid appealing to *properties of limit systems* in accounting for the AB effect, and instead make use of *limit properties* so that we abide by the sound principle. It turns out that such an approach is possible, but before explicating it I wish to review the standard account of the AB effect in a bit more detail.

To begin, the Hamiltonian that allowed Aharonov and Bohm (1959, 486) to derive the interference shift had the following form:[103]

$$H^I = \frac{1}{2m}(\boldsymbol{P} - \frac{q}{c}\boldsymbol{A})^2$$

However, H^I is not a self-adjoint or essentially self-adjoint operator on $C^\infty(\mathbb{R}^3 \setminus S_{in})$ (i.e., it does not have a unique self-adjoint extension on the domain of smooth compactly supported functions on the configuration space $\mathbb{R}^3 \setminus S_{in}$).[104] This means that, as it stands, H^I is not an observable. Furthermore, specific quantum dynamics are given by a unitary group, and we know from Stone's theorem (see Reed and Simon 1980, 266-268, Theorem VIII.8) that the infinitesimal generator of such a unitary group must be a self-adjoint operator. If an operator is not self-adjoint but it is essentially self-adjoint, then we can use its unique self-adjoint extension (namely, its closure) to generate the dynamics. On the other hand, in the case of H^I there is a plethora of self-adjoint extensions (de Oliveira and Pereira 2010), corresponding to different dynamics (i.e., different predictions for scattering experiments), and given by different possible boundary conditions that depend on the interaction between the particle beam and the solenoid. Thus, questions arise: which self-adjoint extension do we use and what is our justification for using a particular self-adjoint extension over another?

On the received view of the AB effect, we declare that the AB effect set-up is approximately similar to an infinite and absolutely impenetrable solenoid, and so it is in virtue of the idealized set-up that we justify choosing boundary conditions in which wave functions in the domain of H^I vanish at the solenoid boundary. Choosing such a boundary condition corresponds to choosing a particular self-adjoint extension of H^I, which I denote and identify with (what we

---

[103] The reason for a shift in notation from H^I_{AB} to H^I will become evident shortly.
[104] $\mathbb{R}^3 \setminus S_{in}$ is the space that arises when one removes the infinite cylinder from $\mathbb{R}^3$.

have been calling the Aharonov-Bohm Hamiltonian) H_{AB}^{I}, and we continue to derive the interference shift as is done in standard textbooks (see Appendix B). However, two problems arise. First, I have argued that such an approach is not cogent because an appeal is being made to a pathological idealization: we cannot just assume that a longer and well-shielded solenoid will behave as, or is approximately similar to, an infinite and absolutely impenetrable solenoid. Second, the (pathologically idealized) assumption of impenetrability only guarantees that, for a continuously differentiable wave function, the electron probability current $j \coloneqq -i(\Psi^{*}\nabla\Psi - \Psi\nabla\Psi^{*})$ must vanish at the solenoid boundary. This can be implemented using *different* boundary conditions including the Dirichlet boundary conditions ($\Psi = 0$) that Aharonov and Bohm (1959) used, Neuman boundary conditions ($\nabla\Psi = 0$), or Robin boundary conditions ($\nabla\Psi = r\Psi, r \in \mathbb{R}$) (de Oliviera and Pereira 2010). Thus, even while appealing to absolute impenetrability we lack a justification for choosing a particular boundary condition, and hence for making use of H_{AB}^{I} in accounting for the AB effect. *The received view of the AB effect is lacking in justification for choosing the main explanatory component*, namely, H_{AB}^{I}.

Accordingly, I am suggesting an *alternative interpretation*. Consider a family of realistic Hamiltonians $\{H_{L,n}\}$ that are well-posed and self-adjoint on $C^{\infty}(\mathbb{R}^{3})$. For instance, consider $H_{L,n} = \frac{1}{2m}(\boldsymbol{P} - \frac{q}{c}\boldsymbol{A_L})^2 + nV(r)$, where $\boldsymbol{P} = -i\hbar\nabla$ is the momentum operator, $\boldsymbol{A_L}$ is a vector potential that depends on the *finite* length L of the solenoid, $V(r)$ corresponds to the *penetrability* of the solenoid such that $V(r) = V$ if $r \leq r_0$, $V(r) = 0$ if $r \geq r_0$, with r_0 the radius of the solenoid, and n is an integer. Then show that such a family converges to H_{AB}^{I} in the limit in which the solenoid length and impenetrability grow. Schematically, we want the following equation to hold:

$$\lim_{n,L\to\infty}\{H_{L,n}\} = H_{AB}^{I} \quad (6.1)$$

Recent work in mathematical physics proves that the above equation indeed does hold. Magni & Valz-Gris (1995) show that the $n \to \infty$ case holds, while de Oliveira & Pereira (2008, 2010, 2011) include both cases $n \to \infty$ and $L \to \infty$, in which it does not matter which limit is taken first. Specifically, de Oliveira & Pereira (2008) show that the limits exist in the strong resolvent sense, and that Dirichlet boundary conditions are singled out if we represent the assumption of infinite impenetrability with step-function potentials $V_{n,L}$ and take the infinite limit $n \to \infty$. de Oliveira & Pereira (2011) extend the result to norm resolvent convergence (wherein spectral and eigenfunctions convergences are guaranteed) in the case of the $n \to \infty$ limit. Moreover, Ballesteros and Weder (2009a, 2009b, 2011) show that such results hold in the context of experiments confirming the AB effect with toroidal magnets (e.g., Caprez *et al*. (2007), Tonomura *et al*. (1982, 1986)). They provide a quantitative error bound for the difference in norm between the exact solution and Aharonov and Bohm's (1959) idealized wave function solution—what is usually called the AB effect *Ansatz*—thereby showing that the Aharonov-Bohm *Ansatz* is a good approximation of the exact solution.[105]

What is important from our perspective is that although it does not matter which boundary condition one uses in order to attain the interference shifts associated with the AB effect, still, de Oliveira & Pereira (2010) show that different boundary conditions correspond to different physics along with different empirically confirmable predictions for scattering experiments. So *there is a real need to justify the appeal to the AB effect idealizations and tell a*

[105] In addition, Babiker *et al*. (1984) and Roy (1980) considered the infinite solenoid length limit, while Kretzschmar (1965) also considered the infinite potential barrier limit. An additional idealization studied by Weisskopf (1961) via a limiting process concerns the fact that the electromagnetic field associated with the solenoid will interact with surrounding particles when being turned on and off.

story about why the AB effect, conventionally defined on a non-simply connected configuration space, has anything to do with the AB effect as it is manifested in the laboratory. My point is that it is exactly the kind of results discussed above that give us this story.

Furthermore, in order to emphasize the *justificatory difference* between the received view and the alternative interpretation that I am proposing it is important to make to remarks. First, $\lim_{n,L\to\infty}\{H_{L,n}\} = H^I_{AB}$ gives a *justification* for using H^I_{AB} (with the boundary conditions chosen by Aharonov and Bohm (1959) in which the wave function vanishes at the solenoid boundary) that *does not make use of an idealized infinite and absolutely impenetrable solenoid*. The philosophical upshot is that we need not appeal to pathological idealizations in accounting for the AB effect. Second, a story can be told about why, in principle, one need not appeal to H^I_{AB} and talk of an idealized multiply connected space in order to account for the AB effect. Specifically, in virtue of $\lim_{n,L\to\infty}\{H_{L,n}\} = H^I_{AB}$, one can be confident that the AB effect interference shift derived from a sufficiently long and impenetrable solenoid represented by $H_{L,n}$ is well approximated by the idealized treatment. But with the $H_{L,n}$ the AB effect is manifested without the requirement that the electron beam never enters the region where the magnetic field is non-zero. Consequently, it is in virtue of the facts that $\lim_{n,L\to\infty}\{H_{L,n}\} = H^I_{AB}$, that H^I_{AB} can be used in order to attain the phase shifts and interference patterns that arise in the AB effect *in the first place*. That is to say, we gain an explanatory story for why a highly idealized model (specifically, H^I_{AB}) is still empirically adequate. See Figure 6.4 for a pictorial summary.

AB Effect Case

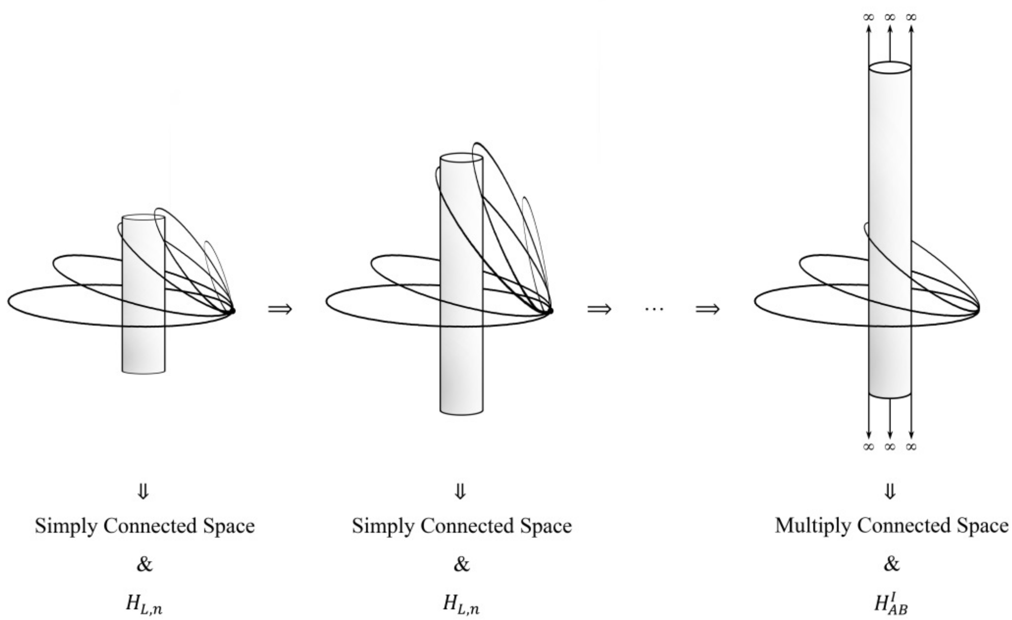


Figure 6.4: The AB effect concerns a sequence of systems of penetrable solenoids in which the length, and the potential barrier representing impermeability, grow to infinity. On the way to the limit, the (configuration) space is simply connected and the corresponding Hamiltonian faithfully representing the system is $H_{L,n}$. At the limit, the (configuration) space is multiply connected, with a corresponding idealized Hamiltonian H^{I}_{AB}.

One may attempt to object, vis-à-vis the fiber bundle formulation of the AB effect, that in such a context an appeal to a multiply connected space truly is necessary in order to account for the effect. The idea is that it follows from the fact that a non-trivial holonomy corresponds to a non-trivial phase factor (giving rise to an interference pattern), that the base space of the principal and associate fiber bundles representing a system in which the AB effect arises must be multiply connected. This is so because if there is a non-trivial holonomy, and if the connection is flat (i.e., the electromagnetic fields are zero), then the base space has a non-trivial topology. In reply, it seems to me that there is a story to tell about how one may attain a non-trivial holonomy without appealing to a multiply connected space. Certainly, if the connection is not flat, this is possible. To that effect, Katanaev (2011, 2012) offers such an alternate (fiber bundle) approach to the AB effect. In other words, there are two options. Either the connection is flat, in which case the base space must be multiply connected for there to be a nontrivial holonomy. Or else, the connection is not flat on some bounded domain, while flat outside. It is this second case that is relevant to the AB effect. Moreover, if the base space representing physical space was truly multiply connected, it is not at all clear that nontrivial holonomies derived in such a manner would have anything to do with the AB effect since the effect is fundamentally dynamical in nature: If we carve out the region of physical space where the solenoid resides, then there is no solenoid, no magnetic field, and no AB effect![106]

In sum, if we think that the AB effect is fundamentally topological in nature, in the sense that we have to appeal to an idealized multiply connected (configuration) space, then we commit ourselves to the claim that a pathological idealization is essential for accounting for the AB effect. This is an untenable view of the necessity of idealizations in science, and in blatant contrast with the intuitions shared by both philosophers and physicist alike as conveyed by the sound principle. For these reasons I urge that we take the alternative view of idealizations discussed in this section. The moral may be stated generally as follows:

---

[106] Thank you to John Earman for making this point crystal clear.

> Hasty idealization often obstructs the application of the theory. The remedy is always the same: to consider the ideal system as the limit of more *realistic* ones, to which the theory can be consistently applied. (Berry 1986, 320; my emphasis)

For context, it is worthwhile to note that Berry (1986) is explaining how it is the limiting procedures of the type I discussed in this section that justify the reality of the AB effect. Without such results, i.e., without appropriate justifications for appealing to idealizations, one may worry that the effect is an artifact of an idealization. In fact, it is exactly this type of worry that led Bocchieri and Loinger (1978) to claim that the AB effect has a "purely mathematical origin," and gave rise to a twenty-five year controversy regarding the reality of the effect (e.g., Bocchieri and Loinger (1978), Bocchieri *et at*. (1979, 1980), Bohm and Hiley (1979), Roy (1980), Kuper (1980)).

6.5 CONCLUSION

In this chapter I have looked at the nature of some pertinent idealizations that arise in standard scientific accounts of the AB effect. I showed how the received view on the subject takes the idealizations in question to be topological. It is claimed that one must appeal to a multiply connected electron configuration space in order to account for the AB effect. In contrast, I have argued that such idealizations are pathological, for the topological properties of limit systems do not match the corresponding topological limit properties. The received view, then, comprises an unattainable position regarding the necessity of idealizations in science for it is in blatant conflict with the sound principle. Accordingly, I have suggested an alternative approach to the idealizations that arise in said effects. One looks to families of possible realistic Hamiltonians that faithfully represent systems in question, and uses limiting procedures in order to attain Hamiltonians of the same form as their idealized counterparts, which have allowed for successful empirical predictions. This corresponds to a match between limit properties and properties of limit systems with respect to the Hamiltonians representing the systems. In the following chapter I extended my conclusions to the case study of fractional quantum statistics allegedly manifested by excited states in the fractional quantum Hall effect.

7.0 TWO APPROACHES TO FRACTIONAL QUANTUM STATISTICS: PATHOLOGICAL IDEALIZATIONS AND THE CURIOUS CASE OF THE ANYON

In this chapter I consider the claim that anyons, i.e., particles with so-called “fractional” or “intermediate” statics, are physically manifested in fractional quantum Hall effect systems. I identify two different approaches to fractional statistics. The standard approach, (what we may call) *the topological approach* to fractional statistics, is precise and well understood. However, I argue that it appeals to pathological idealizations (see Chapters 5 and 6) and thus suggest an alternative approach. The second alternative approach, *the geometric approach*, does not stem from a solid foundation like the topological approach, so it is not completely clear that the particles in questions are obeying fractional quantum statistics as conventionally defined, but it also does not appeal to pathological idealizations. I make some headway in identifying what kind of work must be done in order to develop the foundations of the geometric phase approach.

7.1 FRACTIONAL QUANTUM STATISTICS AND THE CURIOUS CASE OF THE ANYON

Anyons are hypothetical particles that live in a two-dimensional world.[107] They are distinguished from their well-known brethren, bosons and fermions, by the type of quantum statistics they manifest under particle exchange.[108] That is to say, colloquially, according to non-relativistic quantum mechanics (QM) a system of N identical particles is represented by a wave function $\Psi_{(1,2,\ldots,N)}$ (a quantum state) that satisfies both Schrödinger's equation (which governs the dynamics of the system), as well as specific symmetry properties under particle exchange.[109] In particular, QM tells us that if we exchange the positions of two particles, say particles 1 and 2, the states of the original and permuted systems can differ, at most, by a *phase factor* $e^{i\theta}$, where θ is called the *exchange phase*:

---

[107] Also, anyons have "fractional charge," "fractional spin" and can be thought of as point charged vortices (i.e. point particles with both electric charge and magnetic flux).

[108] I set aside the issue of what exactly one means by "statistics" in this context for Sections 7.3-7.4. Roughly, quantum and statistical mechanics tell us that, given a collection of non-interacting indistinguishable particles at thermal equilibrium, there are two possible ways that the collection might occupy a set of available discrete energy states. The expected number of particles in some specific energy state will depend on the type of particles at hand. Bosons manifest a behavior consistent with Bose-Einstein statistics, while fermions distribute themselves according to Fermi-Dirac statistics. Incidentally, there is a deep connection between the type of statistics that particles manifest and their intrinsic angular momentum (their spin). Bosons come with integer spin and fermions with half-integer spin, while anyons have "fractional spin." An original systematic attempt to prove the spin-statistics relation from fundamental consideration was first provided by Wolfgang Pauli (1940). Ever since then, a plethora of various "spin-statistics theorems" attempting to further ground and generalize such results have arisen. See Duck and Sudarshan (1997) and references within. For my purposes, the discussion of spin will only serve to cloud the points that I am attempting to make. For these reasons I will evade a discussion of spin and spin-statistics relations so far as it is possible. That being said, it is worthwhile to note that there are different approaches to grounding the spin-statistics relations, and it is a matter of controversy which approaches work best, and whether such relations arise also in two-dimensions. Relatedly, the explanation and account for whether and why anyons have "fractional spin," and what relation (if any) this has to their statistics, are all matters of controversy.

[109] This idea that the wave function satisfies certain symmetry properties under particle exchange is known as the symmetrization/anti-symmetrization postulate. If two identical particles are permuted, and if the two particles truly are identical in the sense that they share intrinsic properties such as mass and charge, then we expect both the original and permuted system to have the same observable consequences. Same observable consequence, in turn, means that the wave function describing the system can differ at most by a number of modulus 1. See Section 7.3 for an emphasis on the schematic nature of such reasoning and further discussion.

$$\Psi_{(2,1,\ldots,N)} = e^{i\theta}\Psi_{(1,2,\ldots,N)}$$

In a three-dimensional world θ can take on one of two values: $\theta = 0$ for a system of "bosons" with a corresponding phase factor of $+1$ and $\theta = \pi$ for a system of "fermions" with a corresponding phase factor of -1.[110] Bosons and fermions are the two basic types of particles that exist in our world according to QM. However, in two dimensions a third possibility arises. The exchange phase can take *any* range of values.[111] Particles composing such systems are aptly dubbed *anyons*[112] and their quantum statistics are called "fractional statistics" or "intermediate statistics."[113]

As exciting as such a theoretical construct might be, it turns out that one can conjure up experimental situations that are approximately two-dimensional so that anyons and their properties can manifest in the physical world.[114] Such systems are ones in which we observe the fractional quantum Hall effect (FQHE), the discovery and explanation of which won Robert Laughlin, Horst Störmer and Daniel Tsui the Nobel prize in 1998.[115] However, from a philosophical perspective, at least one worrisome issue arises: If, strictly speaking, anyons exist

---

[110] It is also common to define a statistics parameter $\alpha \equiv \frac{\theta}{\pi}$ where $\alpha = 0$ for bosons, $\alpha = 1$ for fermions, and $0 < \alpha < 1$ for anyons (where, in principle, α can be a rational or irrational number).

[111] θ is defined mod 2π.

[112] The name is due to Frank Wilczek in his (1982b). Note that anyons and fractional statistics have nothing to do with so-called paraparticles and parastatistics (which arise from higher dimensional representations of the permutation group).

[113] For standard textbook accounts and introductions to anyons and fractional statistics see Khare (2005), Lerda (1992), Rao (2001), Stern (2008), and Wilczek (1990).

[114] This is especially interesting because (non-abelian) anyons, such as bound states of the Mayorana Fermion, are the best candidates from which to build quantum computers (Nayak *et al*. (2008)). See Mourik *et a*l. (2012) for recent experimental results and Pachos (2012) for more on anyons in the context of quantum computing.

[115] For more on the integer and fractional quantum Hall effects see Chakraborty and Pietilinen (1995), Douçot *et al*. (2004), Ezawa (2013), Prange and Grivin (1987), Stern (2008), and Yoshioka (2002). For a history and introductory overview see von Klitzing (2004).

solely in two-dimensions—that is to say, they are idealizations of sorts akin to frictionless planes, massless test particles, perfect triangles, and the like—how is it that they can manifest in the real world, which is three-dimensional? How can such idealizations play an essential role in accounting for an experimentally observed physical effect such as the FQHE?

My goal is to make some advancements in answering such questions. Specifically, after a short introduction to the classical and quantum Hall effects in Section 7.2, I will outline two approaches to fractional statistics in Sections 7.3 and 7.4. The first approach, *the topological approach*, is the standard account, but I will argue that it appeals to pathological idealizations (as discussed in Chapter 5) and thus ought to be avoided. The second approach, *the geometric approach*, is based on a less firm foundation than the former approach. Nevertheless, it does not seem to appeal to pathological idealizations and so, drawing on an analogy with my account of the AB effect (Chapter 6), I suggest that it is the correct approach that needs to be taken in order to understand fractional statistics as they are manifested in the FQHE.

In a bit more detail, in Section 7.3 I present the topological approach to fractional statistics and anyons, which is grounded in the so-called configuration space framework to permutation invariance in QM. Here I will explain how fractional statistics arise from a topological idealization. In a nutshell, on this approach anyons and fractional statistics emerge by considering the one-dimensional unitary representation of the fundamental group (the first homotopy group) of the configuration space of identical particles in two-dimensions. This will be contrasted, in Section 7.4, with the manner in which fractional statistics arise in the presence of the FQHE, via a calculation of the geometric phase of a doubly permuted excited FQHE state (a quasihole/quasiparticle). The main take-home message will be that, while fractional statistics qua the topological approach concerns a topological idealization that is pathological, the alternative approach that makes use of a geometric phase does not seem to depend essentially on topology (or the dimensionality of the system).

7.2 INTRODUCTION TO THE QUANTUM HALL EFFECTS

7.2.1 The Classical Hall Effect

In 1879, Edwin Hall published a paper titled "On the New Action of the Magnet on Electric Currents" in which he quotes what he took to be a statement "contrary to the most natural supposition:"

> It must be carefully remembered, that the mechanical force which urges a conductor carrying a current across the lines of magnetic force, acts, not on the electric current, but on the conductor which carries it… if a current itself be free to choose any path through a fixed conductor or network of wires, then when a constant magnetic force is made to act on the system, the path of the current through the conductors is not permanently altered, but after certain transient phenomena, called induction currents, have subsided, the distribution of the current will be found to be the same as if no magnetic force were in action. The only force which acts on the electric currents is the electromotive force… [as quoted in (Hall 1879, 287)]

The quote is taken from Maxwell's *Treatise on Electricity and Magnetism* and, in fact, later editions would bracket this statement and add that "the statement in brackets must be regarded as only approximately true" (Maxwell, 1892, 157). The reason for this includes three sets of experiments conducted by Hall as part of his doctoral studies at Johns Hopkins on Oct. 7-11th, Oct. 28th and Nov. 12th of the same year of 1879.

The first set of experiments, conducted Oct. 7-11th, was aimed at ascertaining whether or not the resistance of a current carrying conductor was affected by the introduction of a magnetic field, at different intensities and directions. Recall, within an intuitive fluid-model of electricity, a voltage difference V, which is an electric potential difference, also known as the electromotive force, acts as a kind of electrical pressure or tension "pushing" the fluid that is the electrical current I through a body such as a conductor.[116] Roughly, holding some conditions such as temperature fixed, the electric current is proportional to the potential—this is known as Ohm's law—and the constant of proportionality between them is defined to be the resistance R of the body through which the current runs:

$$R = \frac{V}{I}$$

Resistance is a quantity that depends on the material and shape of the resisting body in question. The question Hall attempted to answer concerns whether the resistance is also affected by an applied magnetic field. What he found is that "the magnet's action caused no change in the resistance of the coil" (Hall 1879, 289).

Accordingly, with the second set of experiments, Hall investigated whether or not the current would be affected by the magnetic field, such that it would result in some potential difference along the traverse direction of a conductor. However, and this following point is noteworthy, the experiment failed "owing probably to the fact" that the conductor used "had

[116] To be clear, the electric potential, or electromotive force, is neither a force nor potential energy. While the electric force and electric potential energy are measured in Newtons and Joules, respectively, the electric potential is measured in Joules per Coulomb, so it is an energy per unit charge.

considerable thickness" (Hall 1879, 289). Hall repeated the experiment on Oct. 28th, this time with a very *thin* gold leaf and he succeeded to show, by making use of a "Thompson galvanometer," which is a type of current measuring device, that there indeed was a permanent deflection of the galvanometer needle. That is to say, a permanent voltage difference was being detected, the "Hall voltage" V_H, and it is this phenomenon that we call the Hall effect.

Hall followed this with a third set of experiments on Nov. 12th, which were meant as a type of quantitative confirmation of the phenomenon he observed. He was able to show that the product of the applied current I and magnetic field B were proportional to a current picked up by the galvanometer, and thus, via Ohm's law, proportional to the Hall voltage:

$$BI \propto V_H$$

However, Hall emphasized that phenomenon does not arise "under all circumstances." In particular, a ¼ mm thickness of the conducting body (namely, the gold leaf) was enough to ensure that the galvanometer failed to detect any current/Hall voltage.

One could reconstruct the Hall effect in a contemporary form by placing a conductor in the xy-plane in the presence of magnetic field B_z applied perpendicular to the plane, while applying an electric field E_x along the x-axis that induces a current density j_x moving at velocity v_x (also along the x-axis) (See Figure 7.1). The electrons composing the current will be deflected due to the Lorentz force in the y-axis direction. This, in turn, will induce an additional electric field E_y in the y-axis direction, which is the "Hall Effect." In particular, following standard textbook accounts such as Kittle (2005), we start with Newton's second law, which governs the motion of the electrons in our model:

$$\boldsymbol{F} = \frac{d\boldsymbol{p}}{dt} = m\frac{d\boldsymbol{v}}{dt}$$

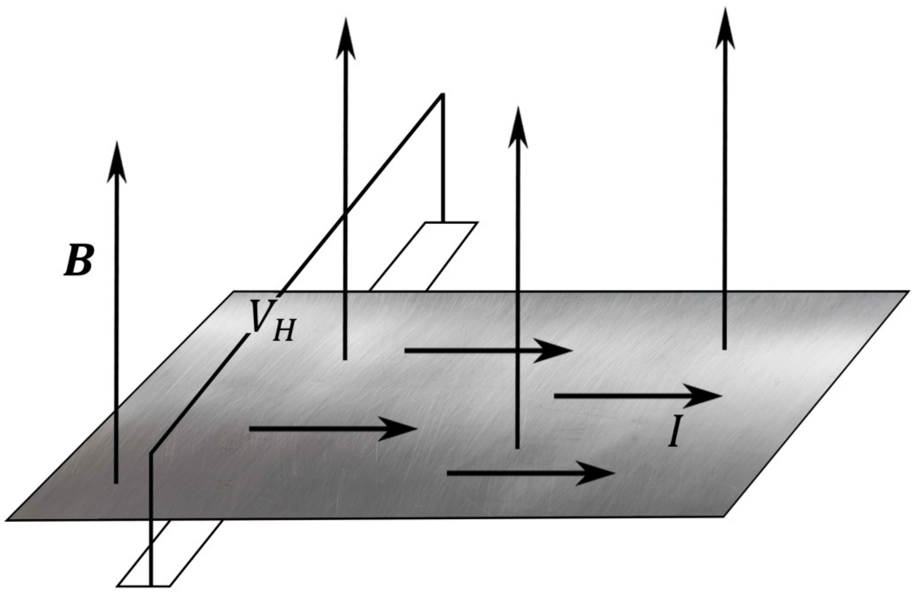


Figure 7.1: The Hall effect. A current i runs through a thin metal conductor in the presence of a uniform magnetic field $\boldsymbol{B}$. There exist a voltage drop V that can be experimentally observed and varies linearly with the magnetic field.

In our case, the force $\boldsymbol{F}$ will be the Lorentz force $\boldsymbol{F_L} = -e(\boldsymbol{E} + \frac{1}{c}\boldsymbol{v} \times \boldsymbol{B})$, where e is the electron charge, $\boldsymbol{E}$ the electric field, c the speed of light constant, $\boldsymbol{v}$ the electron velocity and $\boldsymbol{B}$ the applied magnetic field. However, as for rate of change of the momentum $\boldsymbol{p}$ with respect to time t, we make the following approximation:

$$\frac{1}{m}\frac{d\boldsymbol{p}}{dt} = \frac{d\boldsymbol{v}}{dt} \cong \frac{d\boldsymbol{v_d}}{d\tau}$$

The idea is that the electrons in the conductor are not accelerated by the electric field continually. Rather, they undergo collisions such that the time in between collision is τ and the average velocity between collision is $\boldsymbol{v_d}$, known as the “drift velocity:”

$$v_d \equiv a\tau = \frac{eE\tau}{m}$$

Where a is the acceleration of the electrons (due to the electric field E). The conductance associated with E_y is called the "Hall conductance" and is given by,

$$\sigma_{xy} = \frac{j_x}{E_y} = \frac{nev_x}{v_x B_z} = \frac{ne}{B_z}$$

(where n is the number density). Notice that σ_{xy} varies linearly with ne/B_z. Similarly, through Ohm's law $R = V/I$ we expect the Hall resistance $R_H \coloneqq V_H/I$ to vary linearly with the applied magnetic field. Thus, according to classical electromagnetism, one can see that the Hall conductance (or its inverse quantity the "Hall resistivity") vary *linearly* (*inversely*) with ne/B_z.

7.2.2 The Quantum Hall Effects

Contrary to classical expectation though, approximately a century after Hall's discovery, experiments (von Klitzing *et al*. 1980, Tsui *et al*. 1982) have shown that the Hall resistance is quantized (see Figure 7.2),

$$R_H = \frac{h}{e^2}\frac{1}{\nu}$$

where h is Planck's constant, e the electron charge and the dimensionless number ν—the so-called "filling factor"—has either integer or fractional values. Accordingly, the former case is known as the integral quantum Hall effect (IQHE) the latter case is known as the FQHE.

Systems that manifest the quantum Hall effects make use of semiconductors in strong magnetic fields. These are many-body electron gas systems that are dynamically two-dimensional in the sense that, while the electrons are free to move in two dimensions, they have quantized energy levels in the third dimension.[117] The Hamiltonian for such systems gives rise to discrete eigen-energies called "Landau levels" (LL):

$$E_n = (n + ½)w_c, \ n = 0, 1, 2, \dots$$

Where $w_c := eB/m_e c$ is the cyclotron frequency.[118] These LL are degenerate and it turns out that the filling factor ν is (roughly) the number of electrons per number of states in a Landau level. The IQHE can be understood within the context of non-interacting quantum mechanics in terms of single-particle orbitals of electrons in a magnetic field. Specifically, the IQHE is well explained by LL: as an LL is filled the resistivity increases and once an LL is completely full it takes a finite jump in energy to reach the next LL. The FQHE, on the other hand, is a many-body (electron-electron) effect, and remains mysterious in the absence of further insights.

Robert Laughlin (1983a, 1983b) famously proposed a wavefunction solution to the type of Hamiltonian governing systems in which the quantum Hall effects manifest:[119]

$$\Psi_m = \prod_{j<k}^{N} (z_j - z_k)^m e^{(-\frac{1}{4}\Sigma_l^N |z_l|^2)}$$

---

[117] See Ando *et al*. (1982) for an overview.

[118] The cyclotron frequency is the frequency of a charged particle (in our case, the charge of the electron e) with mass m_e, moving in the presence of a perpendicular and uniform magnetic field with value B. Notice that I have transitioned from using m to signify the mass to the m_e notation, so as not to confuse mass with the odd integer m commonly used in the discussion of the FQHE in the context of fractional filling factors $\nu = \frac{1}{m}$.

[119] Laughlin's wave function was found through a variational method in which one considers an educated guess for the ground state of a known Hamiltonian, dependent on various parameters, and then continues to minimize the energy to get a good approximation for the ground state of the system. The expression that I'm using here for Laughlin's wavefunction is the one that appears in Arovas, Schrieffer, and Wilczek (1984, 282).

where $z_j = x_j + iy_i$ is the complex coordinate denoting the j^{th} particle's position in the plane, and the magnetic length $l_B \coloneqq \sqrt{\hbar c/eB}$ has been set to unity.

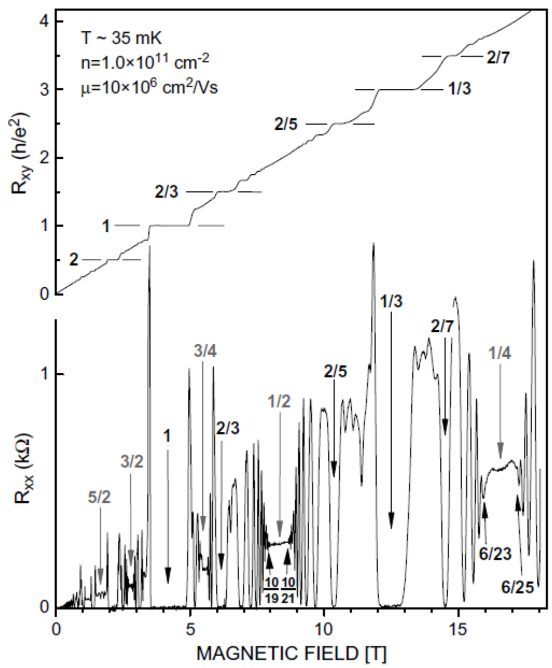


Figure 7.2: Hall resistance $R_H = R_{xy}$ and longitudinal resistance R_{xx} as a function of applied magnetic field. (From Stern 2008, 207)

The wavefunction gave the right description for the fractional filling factors, and Laughlin (1983b) showed that excited states—the quasiparticles and quasiholes—carry fractional charge.[120] In turn, Halperin (1984) and Arovas, Schrieffer and Wilczek (1984) argued that such states obey fractional statistics, i.e., statistics that are neither Bose-Einstein nor Fermi-Dirac statistics, and the quasiparticles (and quasiholes) were identified as anyons based on prior work

[120] The idea is that the fractional values of the filling factor arise from a new state of matter—an incompressible quantum fluid.

by Wilczek (1982a, 1982b).[121] Supposedly, the fractional statistics of FQHE excited states have been confirmed empirically, e.g., Camino *et al*. (2005).[122]

However, as previously discussed, it is worrisome to think that excited FQHE states portray fractional statistics for this is the hallmark of anyons, which are theoretical constructs that can exist only in two dimensions. In order to illustrate this tension we'll want to consider two ways in which fractional statistics may come about. The first characterization of fractional statistics, which we broach in the following Section 7.3, is precise, but will necessarily depend on a two-dimensional idealization. It is in this sense that the standard account of fractional statistics depends on a pathological idealization. Specifically, in analogy with the AB effect (Chapter 6), fractional statistics are thought to be topological effects since it is a topological property of a two-dimensional space—the fundamental group (the first homotopy group)—that will be necessary in order to give rise to fractional statistics.[123] This is similar to the idea that we need to appeal to a multiply connected space in order to account for the AB effect, only in the current context we'll see how we must appeal to the fundamental group of the configuration space of a two-dimensional system of identical particles in order to get fractional statistics.

In contrast, in Section 7.4, we discuss a second manner by which fractional statistics may come about. We will concentrate on the actual calculation conducted by Arovas, Schrieffer and Wilczek (1984), who showed that the excited FQHE states obey such "fractional" statistics.

---

[121] Originally, Laughlin (1983b) postulated that excited FQHE states were bosons and Haldane (1983) thought they were fermions.

[122] That being said, as of yet, the issue is controversial and there is no consensus in the physics community to the effect that fractional statistics actually exist in nature. From my perspective, as will be made clear in Sections 7.3-7.4, this is partially due to a lack of clarity in what is meant by fractional statistics, as well as inadequate foundations for the notion as it manifests in reality.

[123] A mere glance at the literature will confirm that the received view of the nature of idealizations that arises in the FQHE is topological. For example, Khare (2005, 5; original emphasis) states "…in two dimensions, the space is multiply connected which results in the possibility of … *intermediate statistics*" (where intermediate statistics refers to fractional statistics).

Basically, the calculation is one in which we ascertain the geometric phase gained by double permuting the quantum state of two quasiholes,[124] and then interpret this geometric phase as the exchange phase relating to quantum statistics. What is noteworthy about this approach is that it is not at all clear that the two-dimensional idealization is necessary for the manifestation of fractional statistics. However, this leads to a kind of tension between to two approaches because there seems to be a disconnection between fractional statistics qua the first approach—the topological approach—and the fractional statistics qua the second approach based on the geometric phase. The first, but not the second, depends essentially on a topological idealization that is pathological. For this reason it seems to me that there ought to be an account of the fractional statistics manifested by FQHE excited states that does not appeal to the topological considerations of the pathological type. In the following two sections I fill in some details regarding this conjecture in order to support it and solidify my claims.

Before doing so, however, I would like to address a possible point of confusion regarding the two-dimensional nature of the idealization appealed to in the FQHE. In particular, the standard account takes the nature of fractional statistics to be grounded in the kind of topological idealizations that I have discussed, in which the two-dimensional idealization is necessary for fractional statistics to manifest. However, it is conceded that actual FQHE systems portraying fractional statistics are three-dimensional. The claim then is that the systems are "dynamically two-dimensional" in the sense that motion in the third-dimension is quantized. That being said, even dynamically, the systems are not, strictly speaking, two-dimensional. Thus, it will not do to

[124] A similar calculation can be conducted for quasiparticles.

attempt to justify the topological idealization by appeals to the notion of "dynamically two-dimensional."[125] Ando *et al*. (1982, 439) explains:

> These systems are not two-dimensional in a strict sense, both because the wave functions have a finite spatial extent in the third dimension and because electromagnetic fields are not confined to a place but spill out into the third dimension. Theoretical prediction for idealized two-dimensional systems must therefore be modified before they can be compared with experiment.

---

[125] The type of "justification" that I am rejecting is a standard component of the received view (of the nature of idealizations that arise in the FQHE) and is regularly found in canonical textbook accounts of anyons and fractional statistics as can be seen by a glance at a lengthy quote from Khare (2005, 2; emphasis mine):

> The point is that because of the third law of thermodynamics, which states that all the degrees of freedom freeze out in the limit of zero temperature, it is possible to *strictly* confine the electrons to surfaces, or even to lines or points. Thus it may happen that in a strongly confining potential, or at sufficiently low temperature [(both such conditions are satisfied in FQHE experiments)], the excitation energy in one or more direction may be much higher than the average thermal energy of the particles so that those dimensions are *effectively* frozen out. An illustration might be worthwhile here. Consider a two dimensional electron gas… The electrons are confined to the surface of a semiconductor by a strong electric field, and they move more or less freely along the surface. On the other hand, the energy E required to excite motion in the direction perpendicular to the surface is of the order of several milli-electron-Volt (meV). Now at a temperature of say $T = 1K$, the thermal energy is kT, where k is the Boltzmann constant. Thus if the transverse excitation energy is say $10\ meV$, the the fraction of electrons in the lowest excited transverse energy level is
>
> $$e^{-\frac{E}{kT}} = e^{-100} \approx 10^{-44}$$
>
> which is zero for *all practical purposes*. Thus the electron gas is *truly* a two-dimensional gas.

Clearly there is some tension—as well as confusion—regarding whether or not the dynamically two-dimensional system is a two-dimensional system in the strict or approximate sense. Talk of "strictly" and "truly … two-dimensional" seem to confirm the former, while "effectively" and "for all practical purposes" confirm the latter. Certainly, the above calculation will not do as a justification for the idea that FQHE systems are, strictly speaking, dynamically two-dimensional. The fact remains that there is a non-zero probability for an excitation in the third-dimension, and this is all that is needed to cancel out the emergence of anyons vis-à-vis the first approach to fractional statistics (the configuration space approach) appealed to by the received view.

7.3 THE TOPOLOGICAL APPROACH TO FRACTIONAL STATISTICS

I began this chapter by alluding to the idea that there are only two types of particles in nature, bosons and fermions, because this is what is allowed under permutation symmetry. In other words, two types of particles naturally arise from the fact that the wave functions representing the quantum state of a system and its permuted twin can differ at most by a phase factor of modulus one (because the two scenarios are observationally indistinguishable). In fact, this type of reasoning is schematic (at best, and fallacious at worst),[126] and we require a more rigorous framework to permutation invariance in QM that will allow for the emergence of bosons and fermions. There are two such frameworks in the literature. Following Landsman (2013), we will call the first—due to Messiah and Greenberg (1964)—the "operator framework", and the second —due to (among others) Laidlaw and DeWitt (1971), Leinaas and Myrheim (1977), and Wu (1984)—the "configuration space framework".

On the face of it, the two frameworks are different, and Earman (2010) has argued that they have different verdicts regarding the structure of superselection sectors, paraparticles/parastatistics, and permutation symmetry. However, Landsman (2013) shows that in dimensions greater than two ($d > 2$), both approaches can be (made to be) equivalent. That being said, in two dimensions the equivalence fails. The operator framework does not give rise to anyons and fractional statistics, so we shall ignore it here.[127] The configuration space framework, on the other hand, does give rise to anyons and fractional statistics but solely in two dimensions (if we follow the Schrödinger quantization scheme; Heisenberg quantization does not give rise to

[126] See, for example, Dresden (1964) and Kaplan (1994) for an identification of the fallacy.

[127] Still, it is worthwhile to note that if the operator framework turns out to be the correct foundation for permutation invariance in QM, we will lack a corresponding foundation for the notion of fractional statistics.

fractional statistics). In short, on this approach, the type of quantum statistics available depends on an overall phase factor $e^{i\theta} \equiv \gamma$ (for instance, gained by the wave function of a permuted system) which, as it turns out, is the one-dimensional unitary representation of the fundamental group (the first homotopy group π_1) of said system's configuration space. In three dimensions the fundamental group of the configuration space is the (finite and discrete) permutation group S_N which admits of the known one-dimensional unitary representation: $\gamma = \pm 1$ ($+1$ for bosons and -1 for fermions). In two-dimensions, on the other hand, the fundamental group is the (infinite and discrete) braid group B_N with one-dimensional unitary representations giving rise to phase factors of the form: $\gamma_{(\theta)} = e^{i\theta}$where $0 \leq \theta \leq 2\pi$ so that the exchange phase can take on a continuous range of factors allowing for bosons, fermions and anyons. What follows are some details heavily based on Leinaas and Myrheim (1977) and Morandi (1992). The main point is that anyons and fractional statistics, strictly speaking, depend on a topological idealization that is pathological.

7.3.1 The Configuration Space Framework and QM on Multiply Connected Space

Let the configuration space of a particle in d-dimensions be $\mathbb{R}^d$ (where $\mathbb{R}$ is the set of real numbers) so that the position of the particles is given by an element of the space $x \in \mathbb{R}^d$. Consider N such identical particles. The configuration space framework argues that the appropriate configuration space Q for N identical particles is not the Cartesian product of the single particle spaces, $\mathbb{R}^{Nd} \equiv \mathbb{R}^d \times \ldots \times \mathbb{R}^d$ (N times), as one would expect if the particles were distinguishable. Instead, since the particles are indistinguishable, configurations that differ only by a permutation of particles ought to correspond to the same physical state. For the simplest case $d = 1$ $N = 2$, this means that the two configurations $(x_1, x_2) = (1,2)$ and $(x_1, x_2) = (2,1)$ actually represent the same state, and so we must divide out such permuted configurations. In other words, we move from the entire space $\mathbb{R}^{Nd} = \mathbb{R}^{2*1} = \mathbb{R}^2$ of the two-dimensional plane to consider only half the plane (see Figure 7.3).

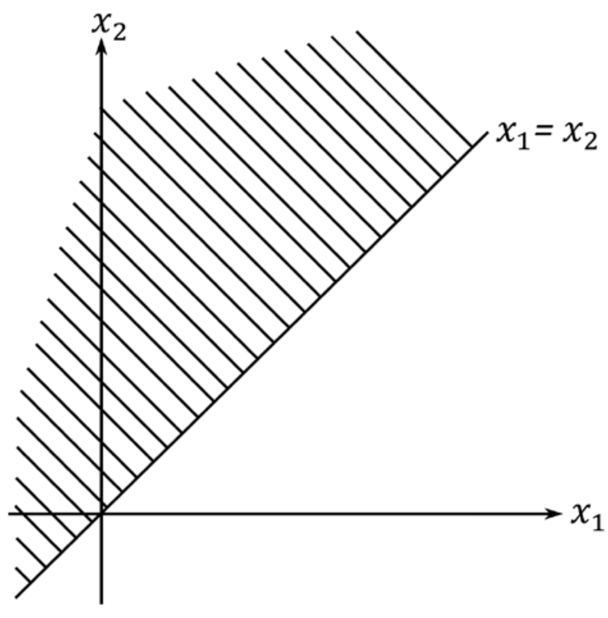


Figure 7.3: The configuration space $Q = \frac{\mathbb{R}^2}{S_2}$ of two identical particles on a line, represented by the non-shaded half-plane.

In the context of the general case, this corresponds to considering the quotient (or identification) space that one attains by diving out the permutation group S_N: $Q = \mathbb{R}^{Nd}/S_N$.

Next, we'll want to excise the set Δ of diagonal points in $\mathbb{R}^{Nd}$, which represents all points where the particles coincide. For the $d = 1$ $N = 2$ case, this means that we must excise points of the sort: $(x_1, x_2) = (x, x)$, where $x_1 = x_2$. If we do not excise coincidence points (known as “singular points”) the resulting configuration space will not have a structure rich enough to represent fermions (or anyons). This is a somewhat ad hoc justification. Nevertheless it is given by some of the pioneers of the configuration space approach.[128] It is sometimes peppered with a statement to the effect that since fermions and anyons satisfy Pauli exclusion-type principles, the removal of coincidence points is justified.[129] Since it is not my intention to advocate one approach to permutation invariance over another, but only to show how the emergences of anyons depend on topological idealization, I'll set the issue aside.

We are left with a configuration space for N identical particles in d dimensions, in which we have divided out the action of the permutation group and excluded diagonal points: $(\mathbb{R}^{Nd} \setminus \Delta)/S_N$. The exclusion of the diagonal point implies that the space in not simply connected and I denote this new configuration space with $\widetilde{Q}$. We must now undertake QM on a multiply connected space. In order to do so I will follow Morandi (1992, 114-144) closely.

Ordinary QM (in the Schrödinger picture) on simply connected regions represents the states of physical systems by complex wave functions Ψ which are elements of a Hilbert space $\mathcal{H}$ of square-integrable functions on the configuration space Q over the field of complex numbers $\mathbb{C}$: $\Psi \in \mathcal{H} = L^2_{\mathbb{C}}(Q)$, where $\mathcal{H}^N = L^2_{\mathbb{C}}(Q) \times \ldots \times L^2_{\mathbb{C}}(Q) \cong L^2_{\mathbb{C}}(Q^N)$ for N identical particles. In order to extend ordinary QM to multiply connected regions we need to appeal to the topological notion of a *universal covering space* Q of $\tilde{Q}$. That is to say, for any topological

[128] For example, Laidlaw and DeWitt (1971, 1377):

> Whether or not two point particles can simultaneously occupy the same point in space is not a question that we wish to settle here. We are only saying that by excluding points of coincidence from the configuration space, the resulting topology leads to meaningful physical results.

[129] For instance, Khare (2005, 23):

> [To] calculate the configuration space of identical particles, such singular points must be excluded… We shall see later that this hardcore constraint is actually unnecessary since for all particles except bosons, there is an automatic angular momentum barrier preventing the crossing of the trajectories. On the other hand, for bosons there is no potential battier and the trajectories may indeed cross but that is fine since, in any case, the relative phase factor for the exchange of two bosons is one!

space, including ones that are not simply connected such as $\tilde{Q}$, we can construct a universal covering space Q that is simply connected (so that ordinary QM applies) with a *covering projection* map: $\pi: Q \to \tilde{Q}$ (where $q \in Q$ and $\tilde{q} \in \tilde{Q}$).

Since all the physical information in QM is contained in the squared modulus of the wave function (i.e., the probability density) $|\Psi(q)|^2$, we will want this quantity to be *projectable* down to $\tilde{Q}$ for any $q \in Q$ in the sense that $|\Psi(q)|^2$ will depend only on a point $\tilde{q} = \pi(q)$ in the multiply connected configuration space $\tilde{Q}$. If this condition is satisfied we say that we have a *projectable quantum mechanics*, and we will have succeeded in extending ordinary QM to a multiply connected region. To that effect, consider an arbitrary point $\tilde{q} \in \tilde{Q}$, let $\tilde{C}$ be a closed curve beginning and ending at $\tilde{q}$, and let $[\tilde{C}]$ be the corresponding (first) homotopy class of curves based at $\tilde{q}$ so that $[\tilde{C}] \in \pi_1(\tilde{Q}, \tilde{q})$.[130] Further, let $q = \pi^{-1}(\tilde{q})$ be any point in the fiber over $\tilde{q}$.[131] The homotopy lifting theorem says that all curves $\tilde{C}$ in $[\tilde{C}]$ are lifted to a curve C in Q beginning at q and ending at some point q' that is also in the fiber over $\tilde{q}$ (so that $\pi(q') = \tilde{q}$ and $q' \in Q$).[132] Denote this result by $q' = [\tilde{C}] \cdot q$ and recall that γ represents a phase factor (although, as of now, we have placed no constraints on the form of γ). We then have the following two central theorems (Morandi 1992, 119-120):

Theorem 1. Projectable Quantum Mechanics are obtained if and only if the wave functions on the universal covering space obey the boundary conditions:

---

[130] First, remember that $\pi_1 \neq \pi$: π_1 represent the fundemantal group (the first homotopy group) and π the covering projection. Second, roughly, recall that any two closed curves (loops) that can be continuously deformed into one another will be part of the same homotopy class. Homotopy is an equivalence relation among loops.

[131] See Appendix C for an overview of the essentials concerning fiber bundles.

[132] C is not necessarily a closed curve. If the bundle is not curved then C is closed but if the bundle is curved then C is not closed.

$$\Psi([\tilde{C}\,]\cdot q) = \gamma([\tilde{C}])\Psi(\mathrm{q}) \text{ for all } q \in Q$$

$$\text{Where } \left|\gamma([\tilde{C}])\right| = 1 \text{ for all } [\tilde{C}] \in \pi_1(\tilde{Q})$$

This means that wave functions on the universal covering space at different points (e.g., q and q'), but on the same fiber above some point in the multiply connected configuration space (e.g., $\tilde{q}$), can differ at most by a phase factor of modulus 1. Moreover, since the universal covering space is simply connected, the wave functions must be single valued. We then get:

Theorem 2. The map $\gamma\colon \pi_1(\tilde{Q}) \to U(1)$ by $[\tilde{C}] \to \gamma([\tilde{C}])$ is a one-dimensional unitary representation of $\pi_1(\tilde{Q})$.

Accordingly, we get our main result: In order to ascertain what type of phase factor is gained by a wave function when it is permuted—with the corresponding available quantum statistics—we must enquire into the one-dimensional unitary representation of the fundamental group of the configuration space of the system. Recall that in the case of N identical particles in d dimensions we have:

$$\tilde{Q} = (\mathbb{R}^{Nd} \setminus \Delta)/S_N$$

It has been shown by Artin (1947), Fadell and Neuwrith (1962), and Fox and Neuwrith (1962) that the fundamental group for the two- and three-dimensional cases are given by:[133]

[133] See Birman and Brendle (2005) for a recent survey.

$$\pi_1\left(\tilde{Q}\right) = B_N \text{ for } d = 2$$

$$\pi_1\left(\tilde{Q}\right) = S_N \text{ for } d = 3$$

Where S_N is the permutation group and B_N is the Braid group. Moreover, as stated previously, the one-dimensional representation of S_N is $\gamma = \pm 1$, $+1$ for bosons and -1 for fermions, while the one-dimensional representation of B_N is $\gamma_{(\theta)} = e^{i\theta}$ where $0 \leq \theta \leq 2\pi$ so that the exchange phase can take on a continuous range of factors allowing for bosons, fermions and anyons.

We see then that the emergence of anyons and fractional statistics depends essentially on the two-dimensional idealization. If we de-idealize, even minutely, our theory does not allow for anyons. This marks a failure of the sound principle and corresponds to a mismatch between a limit property—$\pi_1\left(\tilde{Q}\right) = S_N$ and $\gamma = \pm 1$—and a property of a limit system—$\pi_1\left(\tilde{Q}\right) = B_N$ and $\gamma_{(\theta)} = e^{i\theta}$—in similar manner to the received view of AB effect (i.e., interpreted as topological in nature).

We might be able to get a better sense for why this is so by considering the simplest scenario of two particles $N = 2$ in the $d = 2$ and $d = 3$ cases where: [134]

$$\pi_1\left(\frac{\mathbb{R}^2 \backslash \Delta}{S_2}\right) = \pi_1\left(RP_1\right) = Z \text{ for } d = 2$$

$$\pi_1\left(\frac{\mathbb{R}^3 \backslash \Delta}{S_2}\right) = \pi_1\left(RP_2\right) = Z_2 \text{ for } d = 3$$

[134] Such an illustration arises in standard introductory accounts of fractional statistics such as Khare (2005). Note that Z is the cyclic group of order one, i.e., the infinite group of integers under addition. Z_2 is the cyclic group of order two, i.e., it is the multiplicative group of, say, 1 and -1. RP_1 and RP_2 are the real projective one- and two-dimensional spaces, respectively.

For these purposes I have added an appendix to that effect. The main idea is that by transitioning from three dimensions to two dimensions, we have transitioned from a doubly connected space to an infinitely connected space, and *it is this change in topology that allows for intermediate statistics*.

In short, the two-dimensional setting is essential for producing anyons and fractional statistics within the context of the topological approach to fractional statistics, in the same way that it is essential for producing more than two types of homotopy classes in the simplest case of two particles. This is problematic because it means that we are appealing to a pathological idealization when we derive anyonic statistics. Accordingly, I urge we take a different line on the nature fractional statistics that arise in the FQHE. In the following section I sketch the beginning of such an alternative approach by looking at the actual calculation done by Arovas, Schrieffer, and Wilczek (1984) to show that excited FQHE states indeed obey fractional statistics.

7.4 THE GEOMETRIC APPROACH TO FRACTIONAL STATISTICS

The purpose of this section is to support the idea that we might be able to get a notion of fractional statistics or approximate fractional statistics without appealing to the type of pathological idealization that arises in the context of the configuration space approach. I call this alternative approach to fractional statistics *the geometric approach*.

I will begin by presenting the idea of a geometric phase in order to highlight how such a notion does not depend essentially on the topology or dimensionality of a system.[135] Afterward, I

[135] In order to preempt objections to the effect that the geometric phase itself is topological in nature, I refer the reader to Katanaev (2012) who gives a thoroughly geometrical interpretation of the effect.

present some of the steps taken by Arovas, Schrieffer, and Wilczek (1984) to show that excited FQHE are anyons, with the goal of emphasizing that this is done through a calculation of a geometric phase. The upshot will be that we have good reasons to think that a non-topological and non-pathological account of fractional statistics can be ascertained via the geometric approach. Details can be found in Appendix E at the end of the paper, as well as in literature that I will refer the reader to.

Berry's (1984) original paper concerns a non-dynamical phase factor accompanying cyclic evolutions of non-degenerate quantum systems, but there are similar results for degenerate systems and for a cyclic evolution that is not necessarily adiabatic. The main result that concerns us is that the exchange phase gained by a system traversing in parameter space has two components. One component corresponds to the usual dynamical phase and the second component θ_G is called the *geometric phase* or *Berry's phase*. It can be expressed more generally as a quantity dependent on both the closed curve C in parameter space and the parameters $\boldsymbol{R} = (X, Y, \dots)$:

$$\theta_G(C) = i \oint_C \langle \psi_n(\boldsymbol{R}) | \nabla_R \psi_n(\boldsymbol{R}) \rangle \cdot d\boldsymbol{R} \tag{7.1}$$

Where ∇_R is the gradient with respect to the parameters $\boldsymbol{R} = (X, Y, \dots)$ (and assuming that $\boldsymbol{R}(0) = R(\boldsymbol{t})$ so that C forms a closed curve).

Moreover, as Berry (1984) showed in the fifth section of his paper, we can interpret the AB effect as a particular instance of a geometric phase by representing an AB effect-type system as a charged particle in a box that is transported around an infinite and impenetrable solenoid. The AB effect phase factor arises as the geometric phase gained by the state of the system as it is transported around the solenoid. For our purposes what is worthwhile to note is that there are no assumptions made that depend essentially on the dimensionality (or on an idealized topology) of the system in question. Specifically, one does not expect a geometric phase calculation to alter if there is some small motion of a system in the third dimension. This is key because the derivation of fractional statistics allegedly obeyed by excited FQHE states (quasiholes/quaisiparticles) is calculated by interpreting the geometric phase that arises (when one quasihole/quaisiparticle surrounds another) as the relevant phase factor corresponding to the discussed quantum statistics. On this geometric approach, and in stark contrast to the topological approach discussed in Section 7.3, it does not seem the two-dimensional idealization is necessary for characterizing the notion of fractional quantum statistics. This means that there are good reasons to think that the following conjecture holds: fractional statistics can be characterized in a manner consistent with the sound principle, i.e., without appeal to topological properties of limit systems that do not match the corresponding limit properties.

My purpose in the rest of this section is solely to repeat some of the steps taken by Arovas, Schrieffer, and Wilczek (1984) to derive fractional statistics in order to emphasize the disconnect between this geometric approach and the topological (and pathological) approach discussed in Section 7.3. I refer the reader to Appendix E for details.

Following Laughlin (1983a, 1983b) and Arovas, Schrieffer, and Wilczek's (1984) closely, let us consider a FQHE system with filling factor $\nu = \frac{1}{m}$ where m is an odd integer, and the applied strong magnetic field $\boldsymbol{B}$ is in the *z-axis* direction corresponding to magnetic flux Φ. We'll dub the Hamiltonian governing the system as H^I_{FQHE}.[136] Laughlin's (1983a, 1983b) celebrated wave function (from Section 7.2.2) for the ground state of H^I_{FQHE} can be used to calculate the state function of an excited state, as well as for a system of two excited states

[136] I use the H^I_{FQHE} notation to emphasize that this is the idealized (I) Hamiltonian corresponding to FQHE systems.

(quasiholes) a and b located at positions z_a and z_b (in the plane), respectively, and is represented by

$$\Psi_m^{z_a z_b} = N_{ab} \prod_i (z_i - z_a)(z_i - z_b)\Psi_m \tag{7.2}$$

where Ψ_m is the ground state and N_{ab} is a normalizing factor.

We can determine the quantum statistics associated with exchanging quasiholes a and b by calculating the geometric phase associated with carrying quasihole a adiabatically around a closed loop C, thereby adding time dependence to $z_a = z_a(t)$, and identifying the geometric phase with the exchange phase. The geometric phase θ_G can be calculated by plugging Equation 7.2 into Equation 7.1. It turns out that when quasihole a encircles quasihole b, the new doubly permuted wavefunction $\psi_m'^{z_a z_b}$ gains an extra geometric phase $\Delta\theta_G = 2\pi\nu$. But recall from Section 7.1 that double permutation leads to a general phase factor with an exchange phase θ:

$$\Psi_m'^{z_a z_b} = e^{i2\theta}\Psi_m^{z_a z_b} = e^{i2\pi\alpha}\Psi_m^{z_a z_b} = e^{i2\pi\nu}\Psi_m^{z_a z_b}$$

Where we have introduced the "statistical parameter" defined as $\alpha \equiv \frac{\theta}{\pi}$. We see that $\alpha = \nu$ and recalling that $\nu = \frac{1}{m}$ where m is an odd integer, it follows that $\theta = \frac{\pi}{m}$. For the m=1 case, $\theta = \pi$ corresponding to Fermi-Dirac statistics. But for other values of m, θ corresponds to anyonic statistics.

Let us recap. In the context of the AB effect we saw in Chapter 6 that one is able to start with a realistic Hamiltonian $H_{L,n}$ that faithfully represents an AB effect scenario, and then consider a family of such Hamiltonians $\{H_{L,n}\}$ in order to show via a limiting procedure $\lim_{n,L\to\infty}\{H_{L,n}\}$ that we arrive at the AB effect Hamiltonian H_{AB}^I, which can account for experimental outcomes. That is to say, by solving Schrödinger's equation with H_{AB}^I we attain the state function Ψ from which the relative phase factor $e^{i\theta}$ corresponding to observable interference patters can be extracted through a calculation of the kind conducted in Appendix B. We note that is in virtue of the fact that the limiting procedure $\lim_{n,L\to\infty}\{H_{L,n}\} = H_{AB}^{I}$ holds, that we can use H_{AB}^{I} in accounting for the AB effect interference patterns. We summarize this information schematically in the following:

$$\lim_{n,L\to\infty}\{H_{L,n}\} = H_{AB}^{I} \Longrightarrow \Psi \Longrightarrow e^{i\theta}$$

Analogously, I gather that it is possible, in principle, to start with a realistic three-dimensional Hamiltonian $H(\boldsymbol{L})$ that faithfully represents FQHE systems and depends on some set of parameters $\boldsymbol{L} = (L_1, L_2, \ldots)$ such that the following holds: by considering a limiting procedure $\lim_{\boldsymbol{L}\to\infty}\{H(\boldsymbol{L})\}$ we can attain a FQHE Hamiltonion that has the same form of Laughlin's (1983a, 1983b) idealized two-dimensional Hamiltonian H_{FQHE}^{I}. Once this is done, the rest of the calculation that gives rise to fractional statistics (via a geometric phase as in this section) should follow as before. Specifically, we use variational methods to find the ground state Ψ_m of H_{FQHE}^{I}, and the corresponding state function of two quasiholes $\Psi_m^{z_a z_b}$. We then calculate the extra geometric phase $\Delta\theta_G$ incurred by double permuting $\Psi_m^{z_a z_b}$, and identify the geometric phase with the exchange phase giving rise to fractional statistics. In short:

$$\lim_{\boldsymbol{L}\to\infty}\{H(\boldsymbol{L})\} = H_{FQHE}^{I} \Longrightarrow \Psi_m \Longrightarrow e^{i\Delta\theta_G}$$

Moreover, this process is done without any appeal to the type of pathological idealizations necessary to make sense of the notion of fractional statistics that arises from the configuration space approach. See Figure 7.4 for a pictorial representation of these claims, where the idea is that there is always a mismatch between a limit property and the property of a limit system with respect to the topology of AB effect and FQHE systems, but not with respect the Hamiltonians governing the system.

Whether my conjecture is correct or not, what is clear is that we need to further elucidate the relation between the notion of fractional statistics via the configuration space approach which depends on topology, on the one hand, and the informal geometric phase calculation given by Avoras *et al*. (1984) which depends on the metric in a Hilbert space of quantum state vectors, on the other hand. I submit that this is an issue that necessitates further study and clarification, and ought to have clear consequences for the philosophical understanding of idealizations in science.

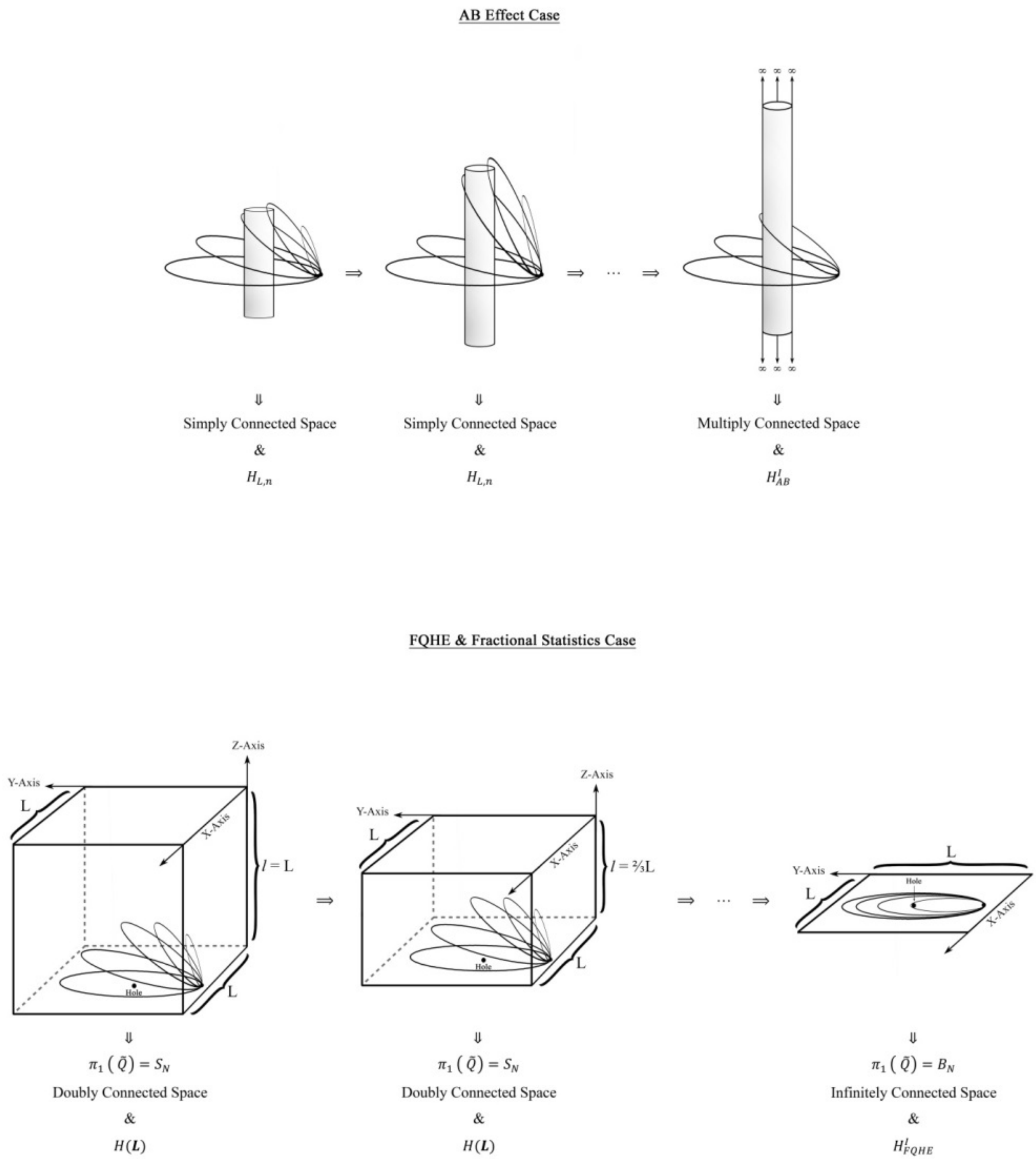


Figure 7.4: Comparison between AB effect and FQHE cases. In the context of the AB effect, if we consider a sequence of solenoids that become longer, we note that as long as we are working

with finite solenoids with corresponding Hamiltonians $H_{L,n}$, then we are dealing with a simply connected topological space. At the limit, in which the solenoid is infinite with a corresponding idealized Hamiltonian H^{I}_{AB}, the topology of the space is multiply connected. We have a mismatch between a limit property and the property of a limit system. Analogously, in the context of the FQHE, we first let a three-dimensional cube represent the configuration space $\tilde{Q}$ of N identical particles in three dimensions (in FQHE conditions), we let a two-dimensional square represent the configuration space $\tilde{Q}$ of N identical particles in two dimensions (in FQHE conditions), and we represent the exclusion of coincidence points (singular points) by a hole. We then consider a sequence of cuboids in which the height becomes shorter and note that as long as we are working with finite heights (i.e. with cuboids) with corresponding Hamiltonians $H(\boldsymbol{L})$, then the topology of the space will be doubly connected and the fundamental group of the configuration space will be the permutation group: $\pi_1\left(\tilde{Q}\right) = S_N$. At the limit, in which the height is zero and we are dealing with a two-dimensional square with a corresponding idealized Hamiltonian H^{I}_{FQHE}, the topology of the space will be infinitely connected and the fundamental group of the configuration space will be the Braid group: $\pi_1\left(\tilde{Q}\right) = B_N$.

7.5 CONCLUSION

Ever since their discovery, the Quantum Hall Effects have gained considerable attention from the scientific community averaging at about one publication per day as of 1995 (von Klitzing 2004). However, to my best knowledge, both the IQHE and FQHE have been largely ignored by the philosophical community.[137] To that effect, I hope to have partially filled in this gap by calling attention to notions such as anyons and fractional statistics, and host of philosophical issues that arise. Accordingly, what I would like to do in this section is briefly discuss some of the issues that merit further study, and then end with a short summary

To begin, there is a wide agreement among (solid state and condensed matter) physicists that the FQHE brings with it various *emergent* phenomena, including fractional charge, fractional statistics, p-wave pairing, chiral edge states, and topological order (wherein attempts at empirically confirming these theoretical structures constitute the forefront of experimental solid state/condensed matter physics). Correspondingly, the philosophical literature on emergence and reduction might greatly gain from attending to these phenomena. Certainly, said phenomena emerge as novel and robust behavior (in reference to some comparison class), and so satisfy the general notion of emergence as articulated by Butterfield (2011). Also, while I have concentrated on the issue of fractional statistics, my discussion exemplifies (at best) only the tip of the iceberg. What is clear from Section 7.3-7.4 is that there are different notions of, and different approaches to, fractional statistics at play in the literature.[138] Although this point has been recently acknowledged,[139] there is need to get clear on what the exact foundations of fractional statistics are. This, in turn, is intimately related to what the correct approach to permutation invariance in QM is. My suggestion in this paper is that, without further qualification, the topological approach to fractional statistics, in which one appeals to topological idealizations that are pathological, will most likely not act as a basis for the type of fractional statistics that might arise in the natural world. Relatedly, it may be possible to argue that fractional statistics are a matter

[137] One prominent exception to these states of affairs includes Earman (2010).
[138] Where some go so far as to reject the notion altogether, e.g. Shrivastava (2005).
[139] See Canright and Johnson (1994).

of convention. For instance, in the context of the Chern-Simons gauge theory the phenomenon of "statistical transmutation" makes it is possible to describe non-interacting anyons and their statistics as interacting bosons or fermions with conventional statistics.[140]

In summary, I have looked at two different approaches to fractional statistics. The topological approach — i.e., the standard account — is a well-founded approach. It is claimed that one must appeal to the two-dimensional configuration space of identical particles, which admits of a continuous range of one-dimensional unitary representations of the fundamental group (the Braid group) of the configuration space, in order to account for the fractional statistics obeyed by excited FQHE states in the natural world. However, I have argued that such idealizations are pathological, since the topological properties of limit systems do not match the corresponding topological limit properties. The standard account, then, comprises an unattainable position regarding the necessity of idealizations in science for it is in blatant conflict with the sound principle. Accordingly, I have suggested an alternative geometric phase approach to the idealizations that arise in said effects. One looks to families of realistic Hamiltonians[141] that faithfully represent systems in question, and uses limiting procedures in order to attain Hamiltonians of the same form as their idealized counterparts which have allowed for successful empirical predictions. This corresponds to a match between limit properties and properties of limit systems with respect to the Hamiltonians representing the systems.

[140] See Chen, Wilczek, Witten, and Halperin (1989), Huang (2003), Jackiw (1990) and Jackiw and Templeton (1981).
[141] Or whatever other mathematical structure is relevant for the theory and phenomenon at hand.

8.0 CAN AN IDEALIZATION BE ESSENTIAL?

The main problem that I have been concerned with in this dissertation has been the essential idealization problem (EIP) (Chapter 2). In order to make headway in solving the problem, Chapter 3 concentrated on the representational relation between theory and world in science, while Chapters 4-8 argued for dispensing with pathological idealizations and for embracing the sound principle. However, a worry arises: If we commit to the sound principle and reject pathological idealization, does this mean that all idealizations are dispensable and that idealization plays no substantive role in science? In this chapter I will answer this question in the negative and outline my main positive account.

8.1 INTRODUCTION

The question at the center of the growing literature on "essential idealizations" (EI) is whether idealizations are genuinely necessary for scientific accounts of physical phenomena. A debate has risen between those who embrace EI (e.g., Batterman (2002, 2003, 2005), Batterman and Rice (2014), Bokulich (2008), Ellis (1992), Morrison (2006), Ruetsche (2011), Weisberg

(2013)),[142] and those who abhor them (e.g., Butterfield (2011), Callender (2001), Earman (2004), Menon and Callender (2013), Norton (2012)). Let us dub the former camp the "essentialists" and the latter camp the "dispensabilists."

What exactly is the main source of conflict between essentialists and dispensabilists is itself a question with no easy answer.[143] That said, I have argued in Chapter 2 that the driving force behind the allegedly paradoxical nature of EI is the worry that there is something inconsistent about such a notion insofar as it would seem that one could reduce an even moderately realist view of science to absurdity via an indispensability-type argument: If idealizations are indispensable to our best scientific theories, we ought to be committed to both their existence and their ontological import, i.e., to idealized objects such as fractals, topologically multiply connected configuration space, and non-analytic partition functions. The example of a non-analytic (or discontinuous) partition function (or thermodynamic potentials), discussed in Chapters 2-3, connects with one of the main contexts in which many battles between essentialists and dispensabilists are fought, in particular, scientific accounts of phase transitions and critical phenomena. The worry, then, might be worded as follows: "Real phase transitions cannot exhibit the discontinuities on pain of contradicting the atomic theory of matter, and, were the discontinuities established factually, the atomic theory would fall" (Norton 2012, 228).[144]

The purpose of this chapter is to show that the division between essentialists and dispensabilists is in fact a false dichotomy. To begin, although I have spent the majority of

[142] It is probably more appropriate to place Ruetsche's (2011) position midway between the essentialists and dispensabilists. Thus, my account here is very much an attempt to build and further her intuitions that, indeed, a peaceful coexistence between the two camps can exist.

[143] See Chapters 4 and 5. Claims abound as to the various puzzles that arise. For example, it may be questionable whether fundamental theory can account for the concrete physical phenomena observed in the world and so one concludes that EI (i) show how "the laws of physics lie" (e.g., Cartwright 1983), (ii) mark a failure of intertheoretic and ontological reduction, and act as a sign of emergence (e.g., Batterman 2002, 2003, 2005) and (iii) support an anti-fundamental conception of science (e.g., Morrison 2006).

[144] In this context, failure of the atomic theory of matter is taken as the consequence of a *reductio*.

dissertation arguing in favor of the sound principle, and against pathological idealizations, in what follows I identify substantive roles for idealizations to play in science that do not contradict the sound principle. In particular, I will identify three substantive roles for idealizations in science, including pedagogical and methodological roles in Section 8.2, and an explanatory role in Section 8.3. Furthermore, in Section 8.4 I propose a working characterization of essential idealizations based on insights offered by both camps and, in doing so, I contend that essentialist and dispensabilists views are importantly complementary.[145] My method will be to show that core claims made about EI by some of the main proponents of the debate, including Batterman (2002, 2003, 2005) (Section 8.3), Butterfield (2011) (Section 8.4.1), Norton (2012) and Ruetsche (2011) (Section 8.4.2), can peacefully coexist in one unified characterization of EI. Section 8.5 will touch on some examples not fully explored in this dissertation, and Section 8.6 ends the chapter with a look at the advancement made in solving the EIP.

In a nutshell, what I will be taking from the essentialists is the idea that many of our best scientific theories appeal to idealized limiting procedures in order to produce a mathematical structure that is used for genuine explanatory and representational purposes (by said theories), and I add that even pathological idealizations can have substantive pedagogical and methodological roles. I reject the dispensabilists' claim that idealizations and approximations are just a matter of mathematical convenience. On the other hand, what I take from the dispensabilists is that there is nothing paradoxical or incoherent about such limiting processes. They are fully justified by the fact that idealized structures and systems arise as the limit of more realistic ones. Or, said differently, what I take from the dispensabilists is the importance of

[145] The dialectics of this paper are inspired by Chakravartty (2010) who argues, in the context of the philosophical debate on scientific representation, that informational and functional theories of scientific representation are importantly complementary. See Chapters 1 and 3 for more details.

subscribing to the sound principle and, hence, rejecting pathological idealizations in explanatory contexts. In other words, my rejection of pathological idealizations is compatible with an embrace of so-called asymptotic and minimal model explanation. *The conflict between essentialists and dispensabilists only arises if pathological idealizations are taken to be essential for asymptotic-minimal model explanation*.

Together, my characterization of EI and the substantive roles that idealizations play in science can be seen as my own attempt to solve (or make substantial headway in solving) the EIP posed at the start of this dissertation.

8.2 ROLES FOR DEALIZATIONS IN SCIENCE

If we subscribe to the sound principle, dispense with pathological idealizations, and in doing so also reject the received view on the nature of idealizations that arise in the context of the AB effect (Chapter 6) and the FQHE (Chapter 7), what role is left for idealizations? It seems to me that we may identify at least three different roles for idealizations corresponding to three different goals that one may have.

8.2.1 Pedagogical Role

First, one might be interested simply in understanding and explaining the mechanics of the scientific theory and formalism appealed to in order to account for an effect. For such purposes idealizations, including pathological idealizations, come in handy. For example, in their original 1959 paper, Aharonov and Bohm had an expository challenge, as is clear from the controversy that arose in the physics community regarding the reality of the effect subsequent to their publication.[146] They had to explain to the physics community that in the context of QM, electromagnetic fields seem to be able to have effects in regions from which they are excluded. One way to flesh this point out concretely is to treat the region $\mathcal{S}_{in}$ inhabited by the magnetic field as topologically disconnected from the region $\mathcal{S}_{out}$ where the electron beam traverses. In fact, this is exactly how talk of "multiply connected regions" arises in the quote from Aharonov and Bohm (1959, 490) in Chapter 6. Thus, *an appeal to a topological and pathological idealization in this context is made in order to explain what the AB effect purports to be in the first place*. Without appealing to pathological idealizations in this context there is worry that what is traditionally identified as the AB effect may be mistaken to be a consequence of, say, the interaction between the magnetic field leaking out of the solenoid and the electron wave function. In contrast, the pathological idealization appealed to by Aharonov and Bohm (1959) via H_{AB}^{I} highlights the fact that *the AB effect is a fundamental feature of QM*: even in the highly idealized, possible but non-actual, scenario in which there is no local interaction between the magnetic field and the electron wave function, QM still predicts that the presence of the magnetic field will affect the wave function in an observable manner. As John Earman has subsequently noted, the AB effect is one manner that highlights the stark contrast between the quantum and classical world.

Similarly, when introducing a physical model of anyons it is customary to follow Wilczek (1982a, 1982b) and do so with the so-called flux-tube model. Here an anyon is described by a spinless particle of charge e in the xy-plane orbiting around a very thin and long

---

[146] See Peshkin and Tonomura (1989) and references therein for discussion.

solenoid with magnetic flux Φ, set perpendicular to the plane, in the direction of the z-axis (see Figure 8.1).

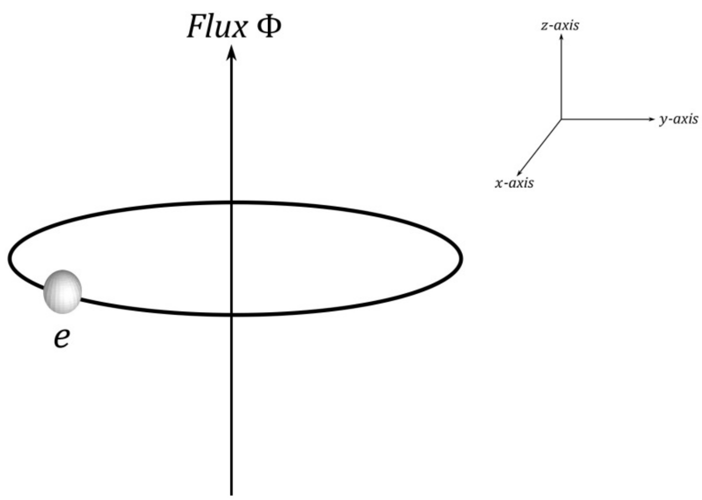


Figure 8.1: The Flux-tube model of the anyon. A spinless charge e particle orbiting around a thin and long solenoid with magnetic flux Φ.

We are then asked to appeal to further idealizations:

> In the limit where the solenoid becomes extremely narrow and the distance between the solenoid and the charged particle is shrunk to zero, the system may be considered as a single composite object — a charged particle-flux tube composite. Furthermore, for a planar system, there can be no extension in the z-direction. Hence, imagine shrinking the solenoid along the z-directions also to a point. The composite object is now point like… (Rao 1992, 15)

Clearly, this highly idealized model serves a *pedagogical* purpose of introducing to the reader the notion of an anyon thought of as a composite particle. But, of course, we must not let too much depend on such idealizations, since it is not at all clear how a very long and extremely narrow solenoid that is shrunk to a point can give rise to any flux whatsoever.

8.2.2 Methodological Role

A different explanatory goal one might have has to do with identifying the different possibilities that may give rise to an effect. As a first *methodological* step in understanding a phenomenon, it seems prudent to attempt to map out various possibilities that might account for the phenomenon. An infinitely long and impenetrable solenoid, corresponding to a configuration space with a non-trivial topology, certainly seems to account for AB effect-type interference patters. Looking to the fiber bundle formulation of the AB effect, if the base space is multiply connected, we are guaranteed to have a non-trivial holonomy. Similarly, fractional statistics can be characterized by appealing to topological properties of a two-dimensional configuration space. These are sensible starting points. That being said, not every possibility is an actuality. The AB effect does *not* arise because of a multiply connected space, and physical manifestations of fractional statistics most likely are *not* consequences of infinitely connected two-dimensional configuration space. There is heuristic, pedagogical and methodological value in mapping out possible accounts of a phenomenon. However, such accounts are far cry from a deep and complete story regarding the scientific account in question.

Rather, the point is that part and parcel of understanding and interpreting scientific theory includes investigating the possible structure of the theory. Recall that in Chapter 3 we discussed that scientific theories do not come with a code or key that tells us what structure in the theory represents physical phenomena in the world. We noted that a code must be deciphered through *empirical investigation* of the world, thereby identifying the target system and phenomenon to be represented and explained. Moreover, it was stressed that *theoretical investigation* is also necessary in order to identify what are the representational capacities and possible structure of our theory. It is absolutely necessary to appeal to pathological idealizations, i.e., highly abstracted limit systems with properties not manifested in the world (not even approximately), in order to fulfill this goal of the scientific endeavor.

Consequently, we may return to our critique of well-received taxonomies of idealization discussed in Chapters 4-5 and note[147] that a *major defect* of the accounts we surveyed (e.g., McMullin (1985), Nowak (1980), Norton (2012), Shaffer (2012), Weisberg (2007a, 2013)) is their overemphasis on idealizations construed as abstract-ideal objects corresponding to some target system, and their neglect of idealization construed as abstract-ideal objects—*that do not correspond to any concrete target system in the world*—and are used to explore the possible structure and representation capacities of our scientific theories.[148] For example, considering the pathological idealization appealed to in the context of the configuration space approach to fractional statistics and Weisberg's (2007a, 2013) three-fold taxonomy of idealization we may ask: Is the two-dimensional idealization a Galilean idealization, a minimalist model idealization,

[147] In accordance with recent discussion with John Earman in the context of the AB effect.

[148] McMullin's (1985) notion of a Construct Idealization is similar to the ideal of an abstract-ideal object that doesn't correspond to a target system in the world. However, he does not stress the methodological *role* that such an idealization plays in science. Similarly, Shaffer's (2012) taxonomy is flexible and robust enough to sustain the idea that some abstract-ideal object to not represent anything concrete in the world, but there is no mention (as far as I can tell) about the need for such objects for the purposes investigating the representational capacities and possible structure of a theory. In the context of the AB effect, for example, Shaffer's (2012) dichotomy between theoretical and non-theoretical idealization is already in tension with the Aharonov and Bohm (1959) treatment since they appeal to both theoretical idealization (in the form of talk of infinite and impenetrable solenoids) and non-theoretical idealizations (in the form of ideal boundary conditions). Thus, it is not clear what the benefit of the taxonomy is in the context of the examples that I have been looking at.

or a multiple model idealization?[149] Since there is no sense in which the fundamental group of the configuration space of identical particles of real systems is the Braid group, not even approximately, it is clear that the idealization is not appealed to only to make some target system computational tractable, so we are not dealing with a Galilean idealization. Similarly, the two-dimensional idealization and Braid group play no causal role in bringing about fractional statistics, and such a pathological idealization cannot play an explanatory role without contradicting the sound principle. Hence, we are not dealing with a minimalist model idealization. Similarly, there are no multiple models in the context of fractional statistics, only one. Talk of multiple models only enters in the alternative geometrical account that I outline, where one looks to a family of realistic Hamiltonians $\{H(\boldsymbol{L})\}$, which faithfully represent FQHE systems, and shows that in the appropriate limit all such Hamiltonians flow to the idealized Laughlin Hamiltonian H^{I}_{FQHE} used to derive fractional statistics. However, in contrast to Weisberg's notion of multiple model idealization, the various Hamiltonians *do not serve different epistemic/pragmatic goals*, *they all serve the same goal* of allowing us to dispense with pathological idealization and justify the original idealized treatment in the first place. Therefore, viewing the story of fractional quantum statistics through the lens of well-received taxonomies of idealization such as Weisberg's (2007a, 2013), seriously distorts and simplifies the state of affairs.

In sum, idealizations and abstractions, including pathological idealizations, are absolutely essential to scientific methodology because they are necessary for mapping out the possible

[149] Recall, from Chapter 4 and Weisberg (2007a, 2013), Galilean idealizations are distortion used to simplify and render computationally tractable the treatment of the target system, minimalist model idealizations are distortions used to expose key causal or explanatory factors in the behavior of the target system, and multiple model idealizations are multiple incomplete models, designed to serve different epistemic/pragmatic goals.

structure and representational capacities of our theories, i.e., they are needed for the purpose of deciphering a code, which is part and parcel of scientific theorizing and practice.

8.3 ASYMPTOTIC EXPLANATION AND MINIMAL MODELS

Over the past two decades Robert Batterman has championed the view that idealizations are indispensible for explanatory purposes and other philosophers have concurred.[150] He dubs his account *asymptotic explanation* or (in the specific case of models) *minimal model explanation*, and it presupposes a distinction between two types of explanatory goals (2002, 23):

> In asking for an explanation of a given phenomenon such as the buckling of a strut, one must be careful to distinguish between two why-questions that might be being asked. … A type (i) why-question asks for an explanation of why a given instance of a pattern is obtained. A type (ii) why-question asks why, in general, patterns of a given type can be expected to obtain.

The idea is that, while a type (i) why-question can be answered with a causal-mechanical or nomological (e.g., covering law) account of explanation, a type (ii) why-question cannot. An answer to a type (ii) why-question necessarily includes an identification of both features that are relevant for the manifestation of some pattern of interest, as well as features that are irrelevant.

[150] E. g., Bokulich (2008), Ellis (1992), Morrison (2006), Ruetsche (2011), Weisberg (2013).

The former features will arise in all instances of the pattern, while the latter will not.[151] Conceptually, all this amounts to idealizing and abstracting away the irrelevant features from particular instances of a given pattern, and then showing how the pattern still obtains.

> Idealization therefore has a fundamental role in physical theory. It is an absolutely necessary device for conceptually isolating the natural processes which are the main subject matter of our inquiries. (Ellis 1992, 265)

More importantly, it is the procedure of "idealizing away" that allows us to identify what are the irrelevant (and relevant) features *to begin with*. Idealizations play a *positive* role by making available *novel* theoretical structures that facilitate answering type (ii) why-questions. In this sense, idealizations—broadly construed—are indispensable.

However, one might reject Batterman's distinction by objecting that type (ii) why-questions are reducible to type (i) why-questions: The reason that patterns of a given type can be expected is because it can be shown that all given instances of the pattern necessarily obtain once we consider the explanation of why a particular instance obtains. For example, according to Hempel and Oppenheim's (1948) revered deductive-nomological ("covering law") account of explanation, a given instance of a pattern obtains because it must necessarily obtain given the corresponding initial conditions and laws of nature relevant to the situation. In the same manner, a pattern obtains because all its instances in fact do obtain, or, said differently, the same effect holds across a wide range of initial conditions.

[151] Assuming that said former features are necessary for a particular instance of a pattern to obtain.

I believe that such an objection is what Norton (2012, 227) has in mind when he claims that the explanation for why "many substances manifest the same critical exponents" is "simply a covering law explanation:"

> Renormalization group methods take the theoretical framework of statistical mechanics as the covering law. They select as the particular conditions a broad class of Hamiltonians pertinent to the material. They then derive universality under conditions close to criticality.

Now, insofar as Norton is sketching a type (i) answer, i.e., a covering law explanation, of why universality obtains for a broad class of Hamiltonians, then he is completely correct. But it does not follow from the above that Batterman's request for a type (ii) explanation is incoherent or unjustified. In fact, on Norton's own account, "the practice of explanation in science is so irregular as to admit no univocal account" (Norton 2012, 227). Relatedly, it certainly seems intuitive that requesting an answer to why an instance of a given pattern obtained is *different* from requesting an answer to why the pattern obtains in general. Furthermore, it is a brute fact that asymptotic methods are used ubiquitously in science for identifying how large classes of systems will behave in some limit regime, where a particular phenomenon of interest might be prevalent, and for recognizing the relative influence of the terms appearing in the equations governing the system. That is to say, a key essentialist insight is that idealizations play an indispensable *explanatory* role, in particular, in accounting for general behaviors and patterns.

This point is emphasized in a more recent work concentrating on models by noting that minimal model explanations provide answers to the following questions (Batterman and Rice 2014, 361):

> Q1. Why are [a set of] common features necessary for the phenomenon to occur?
>
> Q2. Why are the remaining heterogeneous details (those left out of or misrepresented by the model) irrelevant for the occurrence of the phenomenon?
>
> Q3. Why do very different [systems] have features […] in common?

Drawing closely on an analogy with the "renormalization group" (RG) account of "universality" and "critical phenomena,"[152] Batterman and Rice (2014, 362-363) argue that minimal model explanations provide answers to the above questions. Moreover, they claim that Ronald Fisher (1930) offers a minimal-model explanation of the widely observed large-scale pattern of the 1:1 sex ratio in natural populations (Batterman and Rice 2014, 365-373). The general idea is that, first, a space of possible systems is identified. Second, a flow is induced on the space so that irrelevant details corresponding to the phenomenon of interest are eliminated. Last, one searches for "fixed points" induced by the flow in order to identify systems that are in the same "universality class."[153]

What I would like to argue here is that the alternative non-topological interpretation of the AB effect and the fractional statistics that arise in FQHE systems considered in Chapters 6-7 is consistent with asymptotic-minimal model explanations. In particular, an asymptotic-minimal

[152] See Kadanoff (2000).
[153] Notions such as "fixed point" and "universality class" are being used loosely here, and extended beyond their usual scope. One may think of universality in philosophical terms as an instance of multiple realizability. See Batterman (2000).

model explanation of AB effect-type interference patterns is given by showing that the following limiting procedure holds—remembering that I am using the following equation symbolically to represent the the type of work done in, e.g., Ballesteros and Weder (2009a, 2009b, 2011), Magni & Valz-Gris (1995), de Oliveira & Pereira (2008, 2010, 2011):

$$\lim_{n,L\to\infty}\{H_{L,n}\} = H^{I}_{AB} \tag{8.1}$$

In the context of the FQHE this is fleshed out as follows. In order to asymptotically explain how many different FQHE all manifest fractional statistics, and do so without appealing to pathological idealization, we need to show that families of Hamiltonians that realistically represent FQHE converge to the Laughlin Hamiltonian from Section 7.4:

$$\lim_{\boldsymbol{L}\to\infty}\{H(\boldsymbol{L})\} = H^{I}_{FQHE}$$

Concentrating on Batterman's (2002) notion of asymptotic explanation, it is this limiting procedure of Equation 8.1 that offers an explanation of why very different AB effect apparatuses with different levels of solenoid permeability, lengths, radii, vector potentials and magnetic fields, sizes, shapes (cylindrical or torodial), etc., all give rise to similar AB effects. That is to say, we have an explanation for why many different instances—represented by $\{H_{L,n}\}$ (or $\{H(\boldsymbol{L})\}$ in the context of the FQHE)—will give rise to a pattern of a given type that can be theoretically predicted from H^{I}_{AB} (or H^{I}_{FQHE} in the context of the FQHE).

Connecting back with Batterman and Rice's (2014) notion of minimal model explanation, it is the results contained in Equation 8.1 that provide answers to the above questions Q1-Q3. By looking at a family of Hamiltonians $\{H_{L,n}\}$ corresponding to many physical possibilities, and showing that they all "flow" to the same H^{I}_{AB} in the appropriate limit, we are in essence explaining: (Q1) why certain features are necessary for the AB effect to occur, (Q2) why other features are irrelevant, (Q3) and why very different AB effect apparatuses can give rise to an interference shift.[154] In fact, it is the limiting procedure of Equation 8.1 that allows us to recognize that multiply connected electron configuration space is not a relevant feature in the manifestation of AB effect-type interference patterns observed in the world.

This is not to say that the AB effect can be interpreted as an instance of universality.[155] Rather, my claim is that, in the same manner that Batterman and Rice's (2014, 265-274) example from biology, involving sex ratios and populations, is *analogous* to the RG explanation of universality, so is the alternative non-topological account of the AB effect (and of the fractional statistics portrayed by FQHE) discussed in Chapters 6-7. I am arguing that the two explanatory schemes are *analogous*, and that they both provide asymptotic-minimal model explanations—explanations of a structural, as opposed to a causal or covering law, kind. I am not claiming that the alternative account of the AB effect (and fractional statistics) that I discuss in Chapters 6-7 is just an instance of a RG explanation of universality, or that the RG explanatory scheme does not have certain features unique to it.

Still, it may be objected that AB effect-type systems (and FQHE-type systems) are not different enough to admit of an asymptotic-minimal model explanation. Said differently, *there is no interesting type (ii) why-question associated with the AB effect* (or the FQHE).[156] However, I

---

[154] And the same can be similarly said for the FQHE case.
[155] Strictly speaking, the AB effect is not an instance of universality as are, say, the critical exponents associated with continuous phase transitions (see Kadanoff 2000). There is much dissimilarity (e.g., the renormalization group acts on a space of Lagrangians, but I have solely discussed Hamiltonians).
[156] As Robert Batterman has kindly noted.

do not think that this is the case, for an AB effect-type phase factor leading to observable phenomena such as an interference shift, can be observed in widely different systems. For example, quasi-particles such as the kind that arise as excited quantum states of systems portraying the FQHE (as discussed in Chapter 7), manifest exotic quantum statistics in virtue of picking up an AB effect-type phase factor whilst encircling one another. Likewise, there is also reason to think that AB effect type phenomena arise in the context of gravitation (Hohensee *et al*. 2012), and we know that Lorentz covariance implies a similar electric AB effect.

More importantly, once one admits of the distinction between type (i) and type (ii) why-questions, what is the justification for identifying universal phenomena as the kind admitting interesting type (ii) why-questions and rejecting the claim that the AB effect (and/or FQHE) admits of any interesting type (ii) why-question? It seems to me that the story of the AB effect (and the FQHE) calls not only for a causal or DN explanation, but also for a *structural* and thus asymptotic-minimal model explanation, and it is the type of limiting procedure that arises in Equation 8.1 that allows for such an explanation. See Figure 8.2 for a *schematic* pictorial *analogy* between the minimal model explanation of the AB effect, and the case studies discussed in Batterman and Rice (2014).

Another objection[157] that may be raised is that both the limiting procedure and the existence of the limit itself are not key to an asymptotic-minimal model explanation (i.e., a structural explanation). The idea here is that, even without knowing whether the limit in question exists, we can see from the form of the Hamiltonian and the semi-classical approximation used to derive the AB effect, what are the sought after common features needed for the structural explanation to exist for classes of solutions. In reply, it seems to me that even if a structural

[157] As John Earman has kindly noted.

explanation can be given without appealing to any limiting procedures, the process of identifying common features in classes of solutions, many of which either do not or cannot represent any concrete phenomenon, already presupposes appealing to idealizations and abstractions. After all, unless all such solutions correspond to concrete systems in the world, some such solutions can only denote abstract-idealized systems.

Consequently, the conflict between essentialists and dispensabilists is only apparent. On the one hand, the dispensabilists stress that H_{AB}^I alone cannot justifiably account for why a *realistic* system $H_{L,n}$ portrays the AB effect. On the other hand, the essentialist holds that no one realistic system $H_{L,n}$ can explain why *many diverse systems* $\{H_{L,n}\}$ portray an effect that can be derived from an *idealized system* H_{AB}^I. The conflict dissolves once we realize that both essentialists and dispensabilists require appealing to the *same limiting procedures*, of the kind arising in Equation 8.1 ($\lim_{n,L\to\infty}\{H_{L,n}\} = H_{AB}^I$), in order to ground their views: justifying the use of H_{AB}^I, and giving an asymptotic-minimal model explanation for why many different systems in the family $\{H_{L,n}\}$ portray the AB effect (respectively). In principle, a particular $H_{L,n}$ that faithfully represents the states of affairs manifested in an AB effect-apparatus can be used to derive the effect. Thus, idealizations are *not* necessary in order to derive and predict the *occurrence* of the effect. However, the idealized limiting procedures, such as in $\lim_{n,L\to\infty}\{H_{L,n}\}$, are necessary in order to provide explanations to type (ii) why-questions and questions Q1-Q3, as well as to make available the representational structure that plays a role in such explanations.[158] Thus:

[158] Similar statements can be made about the FQHE case.

Idealizations are essential for the explanation and representation of phenomena; they are not essential for the occurrence of phenomena.

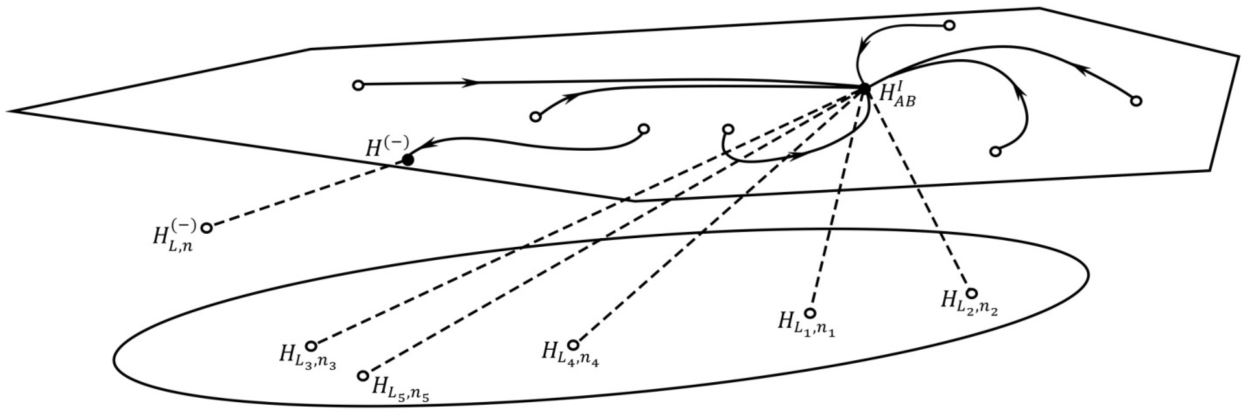


Figure 8.2: The Minimal Model Explanation of the AB effect. The lower collection represents a class of Hamiltonians $\{H_{L,n}\}$, which represent realistic systems, delimited by the fact they all flow to the same Hamiltonian H^I_{AB} under the appropriate limiting procedures (via $\lim_{n,L\to\infty}\{H_{L,n}\} = H^I_{AB}$. $H^{(-)}_{L,n}$ denotes another possible system (e.g., with no magnetic field) that fails to flow to H^I_{AB} , and so is not in the same class. Instead, it flows to some Hamiltonian $H^{(-)}$ that does not give rise to an interference pattern shift.

8.3.1 An Essentialist Objection

An essentialist may object that what I have rejected as a pathological idealization may still play an important role in unifying a host of otherwise unrelated phenomena, e.g., polarization of light, falling cats, parallel parking, Berry's phase (Batterman 2003), and offer an explanation of a holistic and global character (Nounou 2003). Broadly speaking, I am sympathetic to this line of reasoning, but I still hold that *pathological idealizations ought to be avoided*. To that effect, it is important to note that there are two notions of "unification" at play here. For concreteness, I will concentrate on the fiber bundle formulation of the AB effect (for this is the context in which the above claims have been made).

If one means by "unification" (and relatedly by "holism") that we may use the fiber bundle formulation to represent, explain and make predictions about varied phenomena in which non-trivial holonomies arise, then my account accommodates this claim. In the context of the AB effect, we may take Katanaev's (2011, 2012) approach in which the base space is simply connected and the connection is curved on some bounded domain (and flat outside) so that we attain a non-trivial holonomy. In contrast, if one means by "unification" that phenomena with a corresponding fiber bundle formulation in which the base space is not simply connected, are unified with the AB effect by considering its fiber bundle formulation and treating the base space (representing physical space or spacetime) as an idealized multiply connected space, then I reject this line of reasoning. The advantages of explanatory unification are not so great so as to justify "unifying" phenomena on the basis of falsities.

Consider an example discussed in Batterman (2003, 542). The Möbius strip, as a candidate for fiber bundle representation of some physical phenomenon, includes a non-simply connected base space S^1 (a circle) and a flat connection, so that *it is in virtue of the non-simply connected base space* that we can get a non-trivial holonomy (see Figure 8.3). The so-called Hopf (Monopole) bundle,[159] on the other hand, has a simply connected base space S^2 (surface of a sphere) and allows for a non-trivial holonomy *in virtue of a non-flat connection* (see Figure 8.4). Thus, if one means by "unification" that we can use the fiber bundle formalism to talk about

[159] See Batterman (2003, 539-545) for a description of the Hopf bundle and see Wu and Yang (1975) for the original treatment.

non-trivial holonomies, then all three cases (the Möbius strip, the Hopf bundle, and the AB effect) are unified. To that effect, I am in complete agreement with Batterman (2003, 542) who claims the situations are "*strongly analogous*" (his emphasis). But if one means by "unification" that non-trivial holonomies arise in virtue of a non-simply connected base space, then the Hopf bundle and the AB effect are *not* unified with (or analogous to) the Möbius strip in this sense—the non-trivial holonomies appear for *different* reasons.

Batterman (2003, 545) himself distinguishes between two types of anholonomies (i.e., non-trivial holonomies), "topological" anholonomies that arise in virtue of a multiply connected base space and "geometric" anholonomies that appear because of a connection that is not flat. Using this terminology, my claim here is that the AB effect concerns a geometric anholonomy, and that any advantages that one might gain by thinking about the AB effect as (in part) a topological anholonomy will be trumped by the disadvantage of appealing to a pathological idealization (namely, a multiply connected base space).

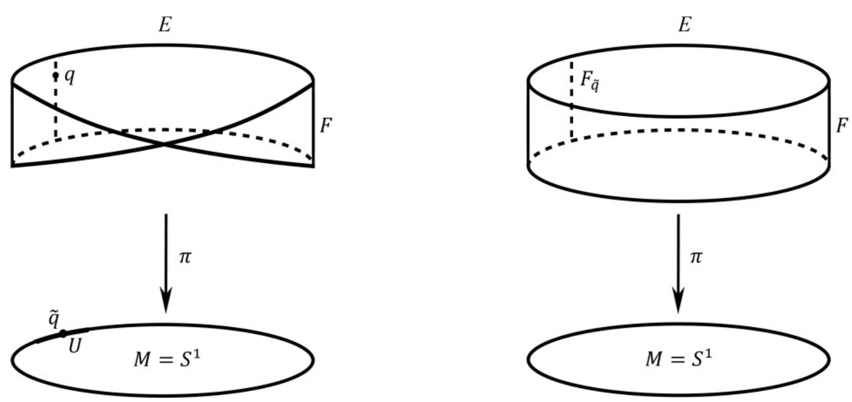


Figure 8.3: Schematic examples of fiber bundles. (Left) A Möbius strip with a non-simply connected base space. (Right) A finite cylinder non-simply connected base space. (See Appendix C for more on fiber bundles.)

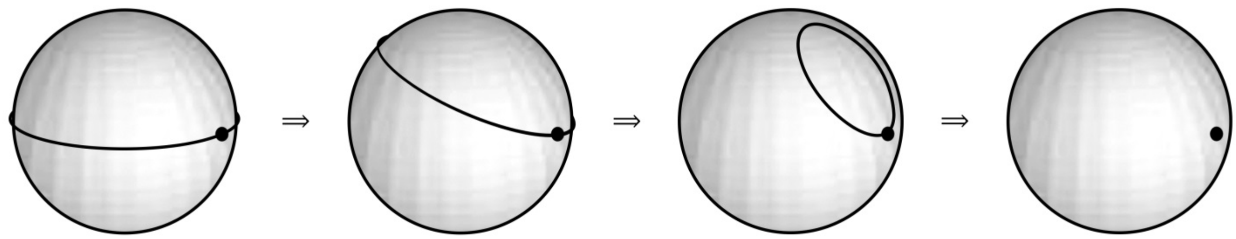


Figure 8.4: The base space of the Hopf (Monopole) bundle is the simply connected surface of a sphere.

8.4 A CHARACTERIZATION

The purpose of this section is to propose a working characterization of (non-pathological) essential idealizations (EI) based on insights offered by both essentialists and dispensabilists, in order to confirm that the idea that the debate between the two camps ultimately revolves around a non-issue and that a more sober midway position is possible.

8.4.1 Emergence or Essential Idealizations?

In a recent paper, Jeremy Butterfield has urged the philosophical community to consider thinking about emergence as "novel and robust behavior relative to some comparison class," which is compatible with reduction—understood as deduction *a la* Nagel (Butterfield 2011, 1065).[160] Two immediate worries arise. The first worry is that Butterfield's notion of emergence fails to account for the historically motivated "central doctrines of emergentism" accepted by the larger philosophical community, such as the causal efficacy and unpredictability of emergent properties (Kim 1999, 20-22). In other words, Butterfield's account misses much of what is interesting to philosophers about emergence. The second worry is that a commitment to traditional Nagelian reduction prohibits Butterfield from appealing to various mathematical methods (e.g., limiting procedures) in "deducing" emergent behavior. In ignoring such prohibitions, Butterfield is avoiding some of the most difficult challenges raised against Nagelian reduction.[161] The main thrust of this worry is that Nagelian reduction is too crude a model to act as a successful account of intertheoretic reduction (although that is not to say that it is not on the right track).[162] In other words, although Butterfield's account of emergence captures some reoccurring features arising within physical theory that are philosophically interesting and merit the consideration of philosophers, it remains an open question whether or not his account illustrates the proper way to think about emergence *per se*. Of course, Butterfield himself is open minded to the idea that emergence might have various characterizations. However, one would worry that without some significant constraints, "emergence" will remain no more than an empty honorific.

My purpose here is not to tackle either of these worries directly. Instead, I want to suggest that by thinking about Butterfield's (2011) alleged examples of emergent phenomena as cases of EI, and by drawing on insights regarding EI from Batterman (2002), Norton (2012), and Ruetsche (2011), we might better focus our attention on the work left to be done in order to

[160] See Nagel (1961) for his account of reduction, and for Nagelian-type accounts see Dizadji-Bahmani *et al*. (2010) and references therein.
[161] For such challenges see Batterman (2010) and Sklar (1993).
[162] See Dizadji-Bahmani *et al*. (2010).

better understand such examples. Accordingly, I offer the following tentative characterization of EI based on Butterfield's (2011) notion of emergence:

> Essential Idealization (first attempt) — Novel and robust mathematical structure that arises on the way to and at some idealizing limit, and secludes those features that are relevant for soundly representing phenomena of interest.

Talk of "on the way to" and "at the limit" is meant to be used in the same sense as Butterfield (2011, Section 3) and will be clarified with terminology from Chapter 5 in the following subsection, while the idea that EI seclude relevant features is elaborated on in following section. What I add here are further observations made in the previous chapters regarding the representational role of EI. The idea is that an EI "soundly represents" in the sense that it allows agents to make valid and sound inferences about the target of representation, but it does not necessarily tells us what the target is like.

We may further amend our characterization of EI by taking into account key essentialists' insight gained in the last section regarding the need of EI in the context of structural, or asymptotic-minimal model, explanations:

> Essential Idealization (second attempt) — Novel and robust mathematical structure that arises on the way to and at some idealizing limit, and secludes those features that are relevant for soundly representing and/or asymptotically explaining phenomena of interest.

8.4.2 Idealizations, Abstractions, or Approximations?

In this section I would like to first consider whether EI are idealizations, abstractions or approximations, and then continue to inquire into the relevant sense of "essentialness." With respect to the first goal, Chapter 4 showed that the main accounts and taxonomies of idealizations largely ignore examples of EI. Instead, we must appeal to recent characterizations suggested by John Norton (2012) and that are discussed in Chapter 5. To see this, recall that although there is no unique accepted characterization of any of the three terms, there is a rough agreement on the following: "idealizations" are meant to introduce simplifying distortions or fictions into some scientific account, while "abstractions" introduce no distortions but instead ignore or abstract away various features of an account. "Approximations" are meant to be purely formal mathematical methods, which are justified within some given context, but do not pose any real philosophical problems.

However, such notions seem irrelevant to the type of case studies arising in the EI literature (such as the example of phase transitions and critical phenomena). For one, the limit in EI certainly does introduce a distortion, so it is not an abstraction, but this distortion is not necessarily simplifying. Rather, the limit is taken to produce the mathematical structure (e.g., non-analytic partition function) which is then used for representational and explanatory purposes. Similarly, although the limiting procedure itself might be taken to be a purely mathematical procedure (and indeed it is a context dependent question whether some system or property and its limit are "approximately" equal), it does not follow that no philosophically interesting issues arise. We noted that, at the very least, that there are important methodological and pedagogical roles for abstract mathematical structure to play, whether one dubs such structure an idealization, abstraction, or an approximation. Thus, although we need characterizations of the notions of “idealization,” “approximation,” etc., in order to assess their putative indispensable role, I do not think that accounts discussed in Chapter 4 fit well with EI.

Instead, it seems then that the best way to think of EI is in terms of Norton’s (2012) recent distinction between an idealization and an approximation, for such a distinction appeals specifically to the type of limiting procedure arising in the context of EI. Recall from Chapter 5 that Norton (2012, Section 3) distinguishes between a system and its limit (i.e. a limit of a sequence of systems)—what he calls a “limit system”—and the property of a system and its limit (i.e. a limit of a property of a sequence of systems)—the “limit property.” A *Nortonian idealization* concerns limit systems and a *Nortonian approximation* concerns limit properties.

Thus, an EI is an “idealization” in the sense that it is either a Nortonian idealization nor approximation, since both Nortonian idealizations and approximations are appealed to by working scientists—as Norton (2012, Section 4) identifies in the context of taking the thermodynamic limit. Norton’s (2012, 226) thesis that we ought to “dispense” with idealizations is then *not* a claim in general against essentialists and EI as we have been discussing here. Rather, the idea is that if we can avoid problems associated with an incompatibility between a limit system and an actual concrete target system by appealing to limit properties, then we ought to do so.

> With this uncertainty, prudence indicates that we should dispense with the idealization, for the approximation already tells us what we could learn from the idealization about the target system. Persisting with the idealization merely risks the error of attributing properties to the limit system that it does not bear. (Norton 2012, 226)

Said differently, and allowing for some interpretative and creative freedom, pathological idealizations, in which there is a mismatch between a limit property and a corresponding property of a limit system, need to be avoided when the target of a scientific account (explanation, representation, etc.) is a concrete system in the world. This seems like a completely reasonable stance for the essentialist to adopt.

It is worth noting though, that appealing to Nortonian idealizations could have the added payoff of making novel predictions and guiding the construction of future theories and models. That is to say, Nortonian idealizations can be used to attribute properties to a target system via a limit system. Such properties can be interpreted as novel predictions made by analogy, which can then be confirmed or disconfirmed. In either case, new models can be constructed accordingly. It is true that such a procedure comes with risk. Analogical inference of the sort always comes with an inductive risk (Norton 2011). But if the analogy holds in the sense that it is empirically confirmed and coheres well with a theoretical background, then we know we are on the right track to a future theory. In the case of phase transitions, the reason that successful accounts appealing to the thermodynamic limit do not overturn the atomic theory of matter is not only because we can take the limit as a Nortonian approximation. Rather, the vast amount of evidence supporting the atomic theory of matter, coupled with the fact that phase transitions qua discontinuity and phase transitions qua a-sufficiently-steep-but-smooth-change are virtually observationally indistinguishable (in the context of predictions tested for in the lab), play significant roles. In short, although prudence does dictate we appeal to Nortonian approximations, successful appeals to Nortonian idealizations might pave the way for novel predictions and future theories.[163]

Next, I'd like to transition to the second goal of this section by considering the sense in which EI are "essential." The claim is that the limiting methods appealed to in the context of EI are essential for the production of the novel and robust mathematical structure that is used for representational and (asymptotically) explanatory purposes. But the notion of "essential" here can come in two different flavors. First, it is a claim about the limiting methods used by our best scientific theories. The idea can be illustrated as follows. Canonical accounts of phase transitions characterize them as discontinuities and appeal to the thermodynamic limit to do so, full stop. As philosophers of science, we ought to take our best present science and its practice seriously. In this sense, the thermodynamic limit (Nortonian idealization or approximation) is essential for the representation and explanation of phase transitions.

A second reading takes EI to be essential in producing mathematical structures that will persist in the success of future theories. Moreover, the two senses need not be in conflict. They can mutually support one another. Laura Ruetsche (2011) summarizes the point elegantly with an especially illuminating example:

> Renormalization Group techniques devised for [quantum statistical mechanics (QSM)] have found application to physicists' [quantum field theory (QFT)]. In QSM those techniques help explain the insensitivity of critical phenomena to the detailed

[163] Of course, with respect to a future theory, what was taken to be an idealization may then be taken to be a faithful representation of the states of affairs. For example, while phase transitions are faithfully represented by a finite-dimensional state space in the context of the atomic theory of matter, on field theoretic accounts of matter such representations are in fact misrepresentations. Similarly, one may argue that if we were to take quantum field theory seriously, phase transitions qua discontinuities (that arise in virtue of taking the thermodynamic limit) are no longer idealizations.

> microphysics of the material prone to them. In physicists' QFT, Renormalization Group techniques help explain how the detailed high-energy physics of exact (but unknown) QFTs could have renormalizable effective theories as their low energy limits. Thus, Renormalization Group techniques, validated by taking thermodynamics quite seriously, are instrumental in identifying plausible future directions for QFTs. Theoretical features made available by idealizations are likely to persist in future theories when those features function as guides for theory development. (Ruetsche 2011, 337)

Incorporating such insights with Norton's distinction we get:

> Essential Idealization (third attempt) — Novel and robust mathematical structure that arises via a Nortonian approximation (or idealization), secludes those features that are relevant for soundly representing and/or asymptotically explaining phenomena of interest, and is essential for the success of present science and/or will underlie the empirical success of future theories.

Last, we may add, in line with Chapters 5-7 and the previous sections in this chapter, that *pathological idealization*, which may play substantive methodological and pedagogical roles, ought to be avoided in the context of representing and explaining concrete target systems in the world.

8.5 CASE STUDIES

In this section I want to emphasize that the conflict between essentialists and dispensabilists is only an apparent one, and illustrate how the working characterization of EI given above (third attempt) acts as a unified account. Constraints of space will allow for a limited number of examples but I submit that what follows can be applied to a wider class of case studies.

8.5.1 The Method of Arbitrary Functions

To start, let us consider one of Butterfield's (2011, 1081-1089) own examples, which concerns the method of arbitrary functions. In particular, we are interested in a theorem by Poincaré which, roughly, states that in the context of a spinning roulette wheel divided into N equal arcs of Red and Black, we can rightfully expect the long-run frequency of Red and Black to be approximately 50%. That is to say, Poincaré showed that if (X, μ) is a sample space partitioned into two subsets of R and B:

> For any $M \in \mathbb{R}$, for all density functions f with derivative bounded by M, $|f'| < M$, as $N =$ number of arcs goes to infinity: $\int_R f d\mu \equiv prob(Red) \to \frac{1}{2}$; and $\int_R f d\mu \equiv prob(Black) \to \frac{1}{2}$. (Butterfield 2011, 1084)

The fact that the probabilities of Red and Black are exactly 50%, and that this is true for *any* biasing profile, is taken to show that such probabilities are novel and robust. What I want to stress is that one cannot derive the equiprobability result without appealing to the $N \to \infty$ limit. In this sense the $N \to \infty$ limit is indeed indispensable.

In addition, the robustness property of Poincaré's results is highly informative. This is so because it is by taking the limit that we find what features of the spinning roulette wheel are irrelevant for the purpose of obtaining the equiprobability result. Such irrelevant features include varying initial positions, angular velocities, frictions present, the actual initial probability density function, etc. Similarly, the $N \to \infty$ limit secludes the relevant features, specifically, the number of intervals N (the more intervals, the closer one will get to equiprobability). Moreover, this information gives us an asymptotic explanation of why the general pattern—i.e., that we can expect the long-run frequency of Red and Black to be approximately 50%—obtains for *any* roulette wheel (assuming N is sufficiently large). That is not to say that one cannot, in principle, also give a covering-law type explanation of the situation by considering many different roulette wheels and their initial conditions, and showing that the above pattern must obtain. But the $N \to \infty$ limit is needed for the type of asymptotic explanation that arises through Poicaré's theorem.

In other words, there is no conflict between the dispensabilists and essentialists here. On the one side, dispensabilists stress that the novel and robust behavior (the equiprobability result) can be mathematically derived using limiting procedures that concern a Nortonian approximation. On the other side, essentialists emphasize that it is exactly such a procedure that allows for asymptotic explanations, along with an identification of the features of the roulette wheel that are (and that are not) relevant for faithfully representing the behavior at hand.

8.5.2 Fractals, Phase Transitions, and KMS States

As a second example, consider the nature of fractals as objects that have irregular (i.e. non-integer) dimension, such as the Cantor set, which is, roughly, the set of points in the interval [0,1] that one gets by repeatedly deleting the open middle thirds, and repeating this process *ad infinitum*. If each stage of deletion is labeled N, then notice that the $N \to \infty$ limit arises in the *definition* of a Cantor set, and so the limit will certainly be essential in, at the very least, representing the Cantor set. Moreover, for any finite N, the "finite Cantor set" will have dimensions of 1. In order to derive the irregular dimensions of a true Cantor set, it is necessary to appeal to the $N \to \infty$ limit. Thus, the novel, robust and informative mathematical structure that is the irregular dimensions of fractals, necessitates the idealized $N \to \infty$ limit in the strongest possible sense (Butterfield 2011, 1090-1103).

Another example are the case phase transitions as they arise in quantum statistical mechanics (QSM), where one can calculate the expectation value of any physical quantity with the "density operator," or "density matrix,"$\hat{\rho}$, which represents the equilibrium state of a system. In particular, within the context of the canonical (Helmholtz Free Energy representation) ensemble, $\hat{\rho}$ takes the following form:

$$\hat{\rho} = \frac{e^{-\beta H}}{Tr(e^{-\beta H})}$$

Where, $\beta = {}^{-1}/_{k_B T}$ is the inverse temperature, k_B the Boltzmann constant, H the Hamiltonian associated with the system (in state $\hat{\rho}$), and $\hat{\rho}$ is called the "Gibbs state" of the system. If the Gibbs state exist, it is unique for any given β and does contain enough structure to represent various thermodynamic phases (at some fixed temperature associated with β) as is necessitated for representing phase transitions. However, there is a theorem that states that whenever $\hat{\rho}$ takes the form above (i.e. the density matrix is the canonical equilibrium density matrix called the Gibbs state), it also satisfies the "KMS condition" (with respect to a given β and H) (Emch and Liu 2002, 251), and such states (that satisfy the KMS condition) are called "KMS States." Moreover, *in the TDL*, KMS states are not unique and the required mathematical structure to represent different thermodynamic phases, and phase transitions, emerge. [164] Thus, we see again that even if it is possible to demote the TDL to the status of a Nortonian approximation, it still seems like a matter of absolute necessity for our best theories of phase transitions to work for us to be able to take the limit in the first place.

8.5.3 Spontaneous Symmetry Breaking in Quantum Field theory

A last example concerns the characterization of spontaneous symmetry breaking (SSB) in quantum field theory (QFT). SSB in QFT arises when the vacuum (ground state) state of system involves degeneracy. However, in the context of QFT, if the vacuum state exists, it is unique. Degeneracy of the vacuum state only arises when a state and its image under an automorphism belong to unitarily inequivalent representations (of the algebra of observables). The failure of the vacuum state to exhibit symmetry is implied by the fact that symmetry of the laws of motion in QFT is not unitarily implementable. In other words, for a symmetry (represented by an automorphism of the algebra) to be spontaneously broken it must be the case that it is not unitarily implementable or that some of the fundamental states of the system related by the symmetry generate unitarily inequivalent representations of the algebra of observables. However,

[164] See Emch and Liu (2002, 346-357) and Sewell (1986, Ch. 4; 2002, 113-123) (and references therein) for the precise results.

one must appeal to the infinite volume limit in order to derive non-unitarily implementable symmetries and unitarily inequivalent representations. Thus the infinite volume limit is essential for deriving a novel, robust and informative mathematical structure that is used to characterize SSB in QFT (Earman 2002, 2003b, 2004; Liu and Emch 2005, Ruetsche 2011).

8.6 CONCLUSION: AN OBJECTION AND THE EIP

In this chapter I have suggested that the division between essentialists and dispensabilists is in fact a false dichotomy, and that views put forth by both camps are importantly complementary as can be seen through a unified (working) characterization of EI, and the substantive pedagogical, methodological, and explanatory roles played by idealizations in science. What I wish to do in this section is address a possible objection from the dispensabilist and end by reconnecting with the essential idealization problem (EIP) that motivated my entire study.

8.6.1 A Dispensabilist's Objection

A dispensabilists might argue that, even in light of the account given here, there is no substantial sense in which idealizations are essential for science. The line of reasoning is as follows. A dispensabilists holds the following two methodological claims. (C1) Whenever an idealization seems essential for a scientific account of some physical phenomenon, then one ought to do some research to see whether the phenomenon can be derived without using idealizations. In the cases I discussed (e.g., phase transitions, AB effect) the research program is successful and so in this case the idealizations are *not* indispensable in any interesting sense. Moreover, (C2) if the research program of type (C1) fails then at the current stage of scientific theorizing the idealization is indispensable. This encompasses a kind of compromise position. However, the ground of this compromise is unstable, for the failure of the research program shows that either (i) the effect at issue has been mischaracterized[165] or else (ii) current theory is false.[166]

I have two remarks in reply. First, I do not think that it follows from the fact that one can dispense with a Nortonian idealization, that "the idealizations are not indispensable in any interesting sense." The limiting procedures appealed to in a (C1)-type research program can be interpreted as idealizations of sorts—or Nortonian approximations—which then turn out to be indispensable for asymptotic explanation. Moreover, we noted that even pathological idealizations have methodological and pedagogical roles to play. So even if a (C1)-type research program succeeds, it does not mean that idealizations have been dispensed with or ought to be dispensed. Rather, all that follows is that it is illegitimate to appeal to a pathological idealization.

Second, it is not at all clear to me why above mentioned compromise position is unstable. All scientific theories, accounts, models, etc., are necessarily idealized (and appeal to misrepresentations) in some sense or other since we do not have (and, presumably will not have any time soon) a final scientific theory of everything. This situation is, was, and will be (in the near future) the state of affairs. Thus, on their own, (i) and (ii) do not mark the sign of an unstable compromise.

---

[165] And a correct characterization will allow its derivation in current theory without use of the idealization.

[166] The dispensabilists may continue: Further methodological imperatives follow. Conduct a research program to investigate (i). If this research program fails, conclude that (ii) is the case, and start a new research program to find a better theory that will explain the effect without having to use blatantly false idealizations. Failure to conduct an investigation by mumbling something about how idealizations are good and indispensable is going to hinder scientific progress. At this point, all the essentialist can do is object that actual scientific practice does not follow such a course and the debate is reduced to a discussion of scientific norms.

8.6.2 The Essential Idealization Problem Revisited

In Chapter 2, I characterized the EIP as follows:

> The Essential Idealization Problem (EIP) — We need an account of how our abstract and essentially idealized **scientific representations** correspond to the concrete systems observed in the world, we need a **characterization** of EI, and a **justification** for appealing to EI's, i.e. an explanation of why and which EI's are successful.

I will end this chapter with noting the headway made in solving the problem. First, in Chapter 3 I gave a content-based account of how our "scientific representations correspond to the concrete systems observed in the world:"

> [cont.] A vehicle *V* is an epistemic representation of a target *T* if and only if *V*'s representational (semantic) contents—determined vis-à- vis a representational code that is adopted and deciphered by intentional agents—are about T.

I also noted that in order to make further progress on the issue one must look to how representational content is identified and determined, and how a representational code is deciphered, in the context of particular case studies involving both theoretical and empirical investigation. Deciphering a representational code presupposes appealing to both non-pathological and pathological idealizations in order to map out the possible structure and representational capacities of a theory.

Next, I gave a working characterization of EI that was constructed on insights from both dispensabilists and essentialists:

> Essential Idealization (third attempt) — Novel and robust mathematical structure that arises via a Nortonian approximation (or idealization), secludes those features that are relevant for soundly representing and/or asymptotically explaining phenomena of interest, and is essential for the success of present science and/or will underlie the empirical success of future theories.

My goal was to show that both camps can maintain their core commitments in one unified and complimentary account of EI. Moreover, the characterization proposed can give a realist account for why EI work via the type of limiting procedures encompassed in Nortonian approximations, namely, idealizations that are products of more realistic representations. It seems to me that significant further work to be done includes investigating, and correspondingly amending, whether the characterization of EI given here can account for other celebrated cases of alleged EI not discussed in this dissertation, or else touched on only briefly. Similarly, it remains an open equation whether my characterization of EI can accommodate the motivation and core tenets of other essentialists, e.g., Bokulich (2008), Morrison (2012).

My justification for appealing to EI remains the same as the justification for appealing to any idealization, namely, the sound principle:

> EARMAN-RUETSCHE SOUND PRINCIPLE − If a scientific account (theory, model) uses an idealization to predict or explain an effect which disappears when the idealization

is removed then either the effect is an artifact of the idealization or else (if experiment confirms the effect) the theory is inadequate.

However, it is necessary to add that pathological idealizations ought to be avoided in the context of explanation and representation of target systems in the world. In other words, EI that are not pathological are successful because they conform to the sound principle (albeit via a non-standard de-idealization scheme as discussed in Chapter 5), and EI that are pathological cannot provide successful explanation. On the other hand, my positive thesis is that even pathological idealizations can play substantive methodological and pedagogical roles, as long as such idealizations do not target any systems in the world. Pathological idealizations are essential for deciphering a representational code, as well as identifying and determining representational content, corresponding to our scientific theories and models.

9.0 MATHEMATICAL EXPLANATION, EASY ROAD NOMINALISM, AND NOMINALISTIC SCIENTIFIC REALISM

It has been recently debated whether or not there is a so-called "easy road" to nominalism in the context of the realism/anti-realism debate in the philosophy of mathematics. In this chapter my goal is to argue that previous discussion of essential and pathological idealizations from Chapters 4-8 has concrete implication for this debate. Specifically, I suggest that the standard topological approach to fractional statistics and the received view of the AB effect (discussed in Chapters 6-7) bear negatively on Mary Leng's account of easy road nominalism, thereby indirectly defending Mark Colyvan's claim that there is no easy road to nominalism. In contrast, the alternative approaches to fractional statistics and the AB effect (outlined in Chapters 6-7) make room for easy road nominalism.

9.1 INTRODUCTION

Recently, several papers have discussed the feasibility of the so-called "easy road" to nominalism.[167] The question motivating (what I will call) the easy road nominalism debate is whether or not it is possible to reject platonism about mathematical objects, while committing to

[167] E.g., Colyvan (2010, 2012), Azzouni (2012), Bueno (2012), Leng (2012), Liggins (2012), and Yablo (2012).

some substantial form of scientific realism, but doing so without taking the "hard road" to nominalism.[168] The hard road includes purging our best scientific theories from quantifying over abstract mathematical objects.[169] Said differently, advocates of the hard road to nominalism claim that mathematical objects are not indispensable to our best scientific theories, and then continue to dispense with such objects. They then are not vulnerable to the Quine-Putnam indispensability argument, a main tenet of which is the Quinean ontic thesis, which states that we are committed to the existence of all the entities that we indispensably quantify over in our best scientific theories.[170] The easy road nominalists, on the other hand, accept the thesis that mathematical entities are indispensable to our best scientific theories. Nevertheless, it is claimed that indispensability does not warrant reification (at least in the case of mathematical entities).

Various strategies have been exploited on behalf of the easy road.[171] For instance, Melia (2000) discusses means by which we might quantify over objects but then deny their existence. That said, as Colyvan (2010) argues, such approaches can only succeed if one rejects the idea there are genuine mathematical explanations in science. However, many proponents of the debate agree that this seems unlikely.[172] What I wish to do in this section is to concentrate on

---

[168] What I have in mind here is some form of scientific realism substantial enough to be moved by the Quine-Putnam indispensability argument and the Quinean ontic thesis, with commitments varying between metaphysical, semantic, epistemic and axiological dimensions. Roughly, the metaphysical dimension has to do with commitment to the existence of a mind-independent world and/or objects within it, (possibly) along with their properties and relations. The semantic dimension concerns interpreting the claims of scientific theories literally, as ones satisfying truth conditions. The epistemic dimension regards the empirical (predictive and retrodictive) and explanatory success of science as evidence for the (approximate) truth of the claims of science, so that scientific claims constitute knowledge about the world and objects within it. The axiological dimension takes the aim of science to give approximately true descriptions and faithful representations of the world. See Boyd (1983, 45), Psillos (1999, xix), Niiniluoto (1999, 21) and Chakravartty (2011). In the following section I will discuss a specific proposal for scientific realism, so-called 'nominalistic scientific realism' (Balagaur 1998).

[169] See Field (1980) for such an attempt.

[170] See Putnam (1971) and Quine (1981), and see Colyvan (2001) for an in-depth study of the indispensability argument and a defense.

[171] See Azzouni (2012), Bueno (2012), Leng (2012), Liggins (2012), and Yablo (2012).

[172] Mainly because of numerous examples that do seem to provide genuine mathematical explanation, e.g., Baker (2005, 2009, 2012), Colyvan (2001, 2007, 2010).

easy road nominalism that is moved by the Quine-Putnam indispensability argument, embraces scientific realism, and accepts the idea that there are instances of genuine mathematical explanations in science. Such an approach is put forth by Mary Leng (2010, 2012), but I submit that the discussion here can easily be extended to cover other approaches to easy road nominalism. I will argue that the standard accounts of the AB effect (discussed in Chapter 6) and fractional quantum statistics manifested by anyons in fractional quantum Hall effect (FQHE) systems (discussed in Chapter 7), bears negatively on Leng's proposal to easy road nominalism, and in doing so I indirectly defend Mark Colyvan's (2010, 2012) claim that there is no easy road to nominalism. However, on the alternative accounts outlined and endorsed in Chapters 6-7, Leng's path to easy road nominalism becomes tenable again. The point of this chapter is to suggest that the easy road nominalism debate can greatly benefit from taking seriously case studies of the sort discussed in this dissertation, all broadly subsumed under the umbrella term of "essential idealizations" (Batterman 2002, 2003, 2005; Ruetsche 2011; Chapters 2-8).

In what follows, I first introduce Leng's (2010, 2012) approach to easy road nominalism in Section 9.2 and show how it depends on the idea that physical structure can "approximately instantiate" mathematical structure when a mathematical explanation is at hand. Next, in Section 9.3, I review the case study of the anyons and fractional statistics presented in Chapter 7, and show that there is no sense in which the mathematical structure appealed to in order to explain fractional statistics is "approximately instantiated" in a physical system. Section 9.4 considers a natural nominalistic reply: explanatory approaches that appeal to pathological idealizations ought to be dispensed with, and alternative approaches, which make way for easy road nominalism, embraced. This is where I also refer to the standard and alternative accounts of the AB effect discussed in Chapter 6. I conclude in Section 9.5. My hope is that it will be clear from my discussion of fractional statistics and pathological idealizations that the study of such examples may greatly inform the debate on easy road nominalism and, more generally, the cogency of mathematical explanations.

9.2 LENG'S EASY ROAD NOMINALISM

Mary Leng's (2010, 2012) approach to easy road nominalism takes the form of a defense of nominalistic scientific realism (NSR), which may be characterized as follows:

> The view that the nominalistic content of empirical science—that is, what empirical science entails about the physical world—is true (or mostly true…), while its platonistic content—that is, what it entails 'about' an abstract mathematical realm—is fictional. (Balagaur 1998, 131)

There is a rough consensus that the plausibility of NSR depends on either excising mathematical explanations from science, or else illustrating "how mathematics can be expected to function successfully in explanatory contexts even if mathematical objects are taken to be mere fictions" (Leng 2012, 986). The later route, taken by Leng (2010, 2012), involves showing that mathematical structure plays a representational role (in the mathematical explanation of a physical phenomenon) such that "…our explanations work by displaying the phenomenon as being a consequence of the (approximate) instantiation of that structure in the empirical situation at hand" (Leng 2012, 990-991).

In other words, we can avoid platonism but remain scientific realists (of the NSR bent) by showing how, in some circumstances, an explanation of a physical phenomenon is given in virtue of the fact that the phenomenon follows from an approximately instantiated physical structure that is represented by some corresponding mathematical structure. These types of explanations, called *structural explanations* by Leng, are fundamentally *mathematical* explanations. However, the fact that the physical structure approximately instantiates the mathematical structure allows us to gain a genuine mathematical explanation without reification.

> A structural explanation will explain a phenomenon by showing (a) that the phenomenon occurs in a physical system instantiating a general mathematical structure, and (b) the existence of that phenomenon is a consequence of the structure characterizing axioms once suitably interpreted. (Leng 2012, 989)

As an example, Leng (2012, 989-990) considers the explanation of why it is impossible to construct a square with the same area of a circle (with only a compass and ruler). The reason is that π (and thus $\sqrt{\pi}$) is a transcendental number.[173] The explanation consists of two parts. First, (a) a physical space approximately instantiates the structure of Euclidean space (and distance between points in Euclidean space approximately instantiates the real numbers structure). Second (b), once we define notions like rule/compass constructions and constructible points, we can show that it follows from the axioms of Euclidean geometry that we cannot construct transcendental (real) numbers.

[173] A transcendental number is a number that is not a root of a (non-trivial) polynomial with rational coefficients (where a rational number is any number that can be expressed as a fraction of two integers).

> Thus, the ultimate applicability of this explanation comes, not from the existence as abstract of any of the mathematical objects involved, but rather, from the interpretation of the geometrical axioms as approximately) true about the physical space of paper and pencil drawings… [T]he explanation is explanatory because it shows an empirical phenomenon (our inability to find certain ruler/compass constructions) to result from the general structural features of the situation. It enables us to see the relation between the impossibility of circle squaring and other impossible constructions, and it enables us to see the phenomenon as resulting from a necessity, not just from bad luck or poor construction. (Leng 2012, 990)

In the following section I will discuss how the standard approach to fractional statistics concerns a genuine mathematical explanation, but there is no precise sense in which the mathematical structure is approximately instantiated in the physical world.

9.3 FRACTIONAL STATISTICS AND APPROXIMATE INTANTIATION

Recall, from Chapter 7, that on the standard configuration space approach to fractional statistics, the type of quantum statistics available to a system depends on the phase factor $e^{i\theta}$ (for instance) gained by the wave function of the permuted system, which turns out to be the *one-dimensional unitary representation of the fundamental group* of said system's configuration space. In three dimensions the fundamental group of the configuration space (of N identical particles) is the (finite and discrete) permutation group S_N which admits of the one-dimensional unitary representation discussed above. That is to say, S_N leads to the phase factor $e^{i\theta} = \pm 1$, where we have $+1$ phase factor for bosons corresponding to $\theta = 0$, and -1 phase factor for fermions corresponding to $\theta = \pi$. These are the conventional quantum statistics. In two dimensions, on the other hand, the fundamental group is the (infinite and discrete) braid group B_N with one-dimensional unitary representations corresponding to a continuous range of phase factors $-1 \leq e^{i\theta} \leq +1$, which gives rise to bosons ($\theta = 0$), fermions ($\theta = \pi$), and anyons ($0 < \theta < \pi$).[174]

In other words, *it is in virtue of the fact that the fundamental group* of the configuration space of identical particles in two dimensions *is the braid group* B_N, and not the permutation group S_N, that we can have anyons and fractional statistics. Thus, we have here a case where we have to appeal to abstract mathematical structure, the fundamental group as the braid group B_N, in order to explain a physical phenomenon, specifically, fractional statistics (as they arise in physical FQHE systems). Moreover, the structure is essentially abstract. It is solely in two dimensions that the fundamental group is the braid group. In three dimensions the structure allowing for fractional statistics disappears.

The explanation of fractional quantum statistics is in essence a mathematical and structural explanation, since it is the *topological structure* of the configuration space that allows for fractional statistics. Specifically, connecting with Leng's (2010, 2012) notion of structural explanation, we may say that (a) the configuration space of FQHE systems are approximately two dimensional (since there is little to no motion in the third spatial direction), and that (b) the existence of fractional statistics is a consequence of two facts. (i) The phase factor characterizing quantum statistics is the one-dimensional unitary representation of the fundamental group of a

[174] Morahndi (1992, 114-144) offers an elegant and precise presentation of these points (where the main theorems that I am making use of arise on pages 119-121). See the appendix at the end of the paper.

system’s configuration space and (ii) in two dimensions the fundamental group is a Braid group. However, it is no longer clear that we can dispense with reifying the explanatory mathematical structure. The reason is that, although FQHE systems are approximately two-dimensional and manifest fractional statistics, it is not the case that the mathematical structure is approximately instantiated in the physical world.

The fundamental group of the configuration space of identical particles in approximately two dimensions is the same as that of three dimensions, the permutation group. In order to allow for fractional statistics we need the fundamental group to be the Braid group, and this can only occur in *exactly* two dimensions. However, physical systems are not exactly two dimensional. Thus, it cannot be said that a physical system “approximately instantiates” the structure associated with the Braid group and necessary for fractional statistics.

Said differently, and making use of the terminology and the example with cuboids presented in Chapter 5, we can say that a real physical system approximately instantiates the structure of our idealized limit system since there exist a correspondence between the limit property and the property of the limit system. In such a context, Leng’s (2010, 2012) talk of “approximately instantiated” structure makes sense. In stark contrast, on the standard topological approach to fractional statistics, “approximate instantiation” *fails*, since we are appealing to a pathological idealization. That is to say, it makes no sense to say that a physical system, which is three dimensional, “approximately instantiates” the topological structure corresponding to the fundamental group qua the Braid group. In three dimensions, the fundamental group is the permutation group S_N that leads to conventional statistics of bosons and fermions. It is only in two dimensions that the fundamental group is the Braid group B_N, which makes room also for anyons with their fractional quantum statistics. Since there is a lack of correspondence between the property of the limit system and the limit property, one can no longer talk about the mathematical structure being approximately instantiated by the physical structure. It is then not clear how, on Leng's approach, one could dispense with reification in this context without undertaking the hard road to nominalism.

9.4 NOMINALISTIC REJOINDER

I have argued that the standard topological approach to fractional statistics contains an appeal to mathematical structure that both plays a genuine explanatory role, and also cannot be dispensed with via talk about "approximate instantiation" in the physical world. In this section I wish to consider an intuitive nominalistic reply, namely, that on some alternative approach to fractional statistics, such as the geometric approach outlined in Chapter 7, one can make sense of "approximate instantiation." To be sure, the nominalist would be putting her neck on line here because the account of fractional statistics sketched is part of the explanatory machinery of our best scientific theories. Nevertheless, she may hold her ground and declare: If an alleged explanation of a physical phenomenon includes appeal to an abstract mathematical structure that cannot be approximately instantiated by a corresponding physical structure, then the explanation is defective. Contrary to the standard explanation of phenomena such as fractional statistics, there is no explanation here at all, and mumblings about indispensability will only hinder scientific progress. Scientists ought to conduct research into explaining the phenomenon in question without appealing to such abstract mathematical structure precisely because it is pathologically idealized.

The nominalist can further her cause by drawing an analogy with the Aharonov-Bohm effect (as I do in Chapter 6): Assume the received view of the AB effect is taken at face value.[175] We ask, what explains the AB effect phase shifts that we observe in the laboratory? The answer is that multiply connected fiber bundle base space is what allows for a non-trivial holonomy corresponding to an interference pattern shift. In other words, we have to appeal to abstract mathematical structure, i.e. a multiply connected topological base space, in order to explain a physical phenomenon, specifically, the AB effect. This then is a mathematical explanation of a physical phenomenon. However, since the multiply connected space concerns a pathological idealization, we can no longer appeal to "approximate instantiation" to purge commitment to abstracta. This is so because a topologically simply connected spacetime cannot approximately instantiate the structure associated with a multiply connected topological space (as I argued in Chapters 5-6). Thus, it seems that the received view of the AB effect brings with it a commitment to platonism *a la* Colyvan (2010) (or else a resolution to take the hard road).

On the other hand, on the alternative approach discussed in Chapter 6, one could tell a story of the kind given by Leng (2010, 2012). We can say that physical systems in which the AB effect manifests approximately instantiate the fiber bundle structure associated with a non-trivial holonomy in virtue of a non-flat connection on some bounded region. Thus, the alternative approach to the nature of idealizations in the AB effect, in which one need not appeal to a non-trivial topological base space, accommodates nominalistic scientific realism. Similarly, by embracing the geometric phase approach to fractional statistics (advertised in Chapter 7), we could take the easy road to nominalism in this particular case study.

[175] For clarity's sake, I'll concentrate on the fiber bundle formulation of the received view.

9.5 CONCLUSION

My main goal in this chapter was to show how there are concrete consequences for the realism/anti-realism debate in the philosophy of mathematics in light of what position one takes on the essential idealization issue discussed in this dissertation. I would like to end with some reflections regarding the plausibility of platonism in light of scientific case studies that seem to appeal essentially to idealizations and indispensable mathematical structure.[176] Specifically, recent efforts on part of platonists to reify mathematical structure revolve around what Alan Baker (2005; 2009, 613) has called the Enhanced Indispensability Argument (EIA) for mathematical objects:

> (1) We ought rationally to believe in the existence of any entity that plays an indispensable explanatory role in our best scientific theories.
>
> (2) Mathematical objects play an indispensable explanatory role in science.
>
> (3) Hence, we ought rationally to believe in the existence of mathematical objects.

It is then clear why platonists ought to be drawn to case studies from science in which mathematical structure used for representational and explanatory purposes arise *solely* in an essentially *abstract* realm. Such ostensible examples of "essential idealizations" automatically satisfy the premises of the EIA and thus, generally, seem to support platonism. It seems then that

[176] Such case studies may be found in Batterman (2002, 2003, 2005), Bokulich (2008), Butterfield (2011), Ruetsche (2011) and throughout this dissertation.

nominalistic scientific realists can further their cause by showing that there exist easy road nominalistic manners by which to understand such phenomena.

Accordingly, I suggest that proponents of the debate on easy road nominalism, as well as the literature on mathematical explanation, may greatly benefit from paying attention to examples from science in which appeals to essential idealizations and indispensable mathematical structure arise. The feasibility of easy road nominalism and NSR depends on assessing whether or not the mathematical structure playing an indispensable explanatory role can be said to be approximately instantiated in the physical structure (which gives rise to the phenomenon to be explained). In order to conduct such assessment philosophers must look to specific case studies and to science itself.

10.0 CONCLUSION

In this dissertation I have attempted to make some headway in understanding various issues that arise in the context of scientific representation and idealization, with a concentration on the essential idealization problem (EIP). My method has been to approach the received literature and positions with irenic sensitivity. From this perspective, it seems to me that denying the existence of essential idealizations and misrepresentations, or holding that all idealizations (abstractions, approximations and misrepresentations) that cannot be straightforwardly de-idealized must be excised, is somewhat rash. Setting aside the instrumental uses of idealizations, I think that a more faithful picture of scientific theory, theorizing and practice is one in which we can make room for substantive roles for idealizations, e.g., via pedagogy, methodology (code decipherment, and content identification and determination), explanation, representation, and guide to future theory and model construction.

That being said, the opposite extreme, which supports a wholesale rejection of the sound principle, and attempts to evade the subsequent foundational puzzles that arise by dubbing them as " red herrings," or appealing to ambiguous notions such as "emergence," is untenable. The traditional reductionist and fundamentalist positions have been too successful in the history and philosophy of science to be completely abandoned.

Accordingly, it seems to me that "the middle way," or a "peaceful coexistence" approach, recommended in this dissertation, may be the most fruitful one for both advancing our understanding of science, and attaining progress in philosophy of science. It is a difficult position to defend for one may quickly slide down a slippery slope to one of the extreme position, e.g., essentialist or dispensabilist, reductionist or emergentist, etc. Nevertheless, I believe it is a position that is philosophically plausible, more interesting, and defendable. This dissertation offers such a defense. Where I have failed, I submit that, with further effort and research, a defense can be given, and so I urge that it be done.

APPENDIX A. CHARGED PARTICLES IN A MAGNETIC FIELD

We are interested in understanding where the Hamiltonian H_{AB}^{I} used by Aharonov and Bohm (1959) comes from. To begin, note that in the context of classical statistical mechanics, we may inquire into how a system of particles evolves over time. If we represent the system by its configuration space (i.e., the possible positions of the particles), we may ask: from the various possible paths that a system might take in configuration space, what is the actual path of the system (i.e., the empirically adequate one)? To that effect, the *principle of least action* (a variational principle) tells us that the path that minimizes the system's *action* will be the actual path traversed by the system. Thus, the dynamics of the system can be determined by the equations that minimize the action. These are the so-called Euler-Lagrange equations, which depend on the Lagrangian L (roughly, the kinetic minus potential energy)[177] of the system.

Following standard textbooks,[178] we want to consider a system with charged (but spinless) particles in electromagnetic fields. The Lagrangian L for a particle of mass m and charge q in the presence of electric and magnetic fields $\boldsymbol{E}$ and $\boldsymbol{B}$ (respectively) is as follows:

$$L(\boldsymbol{r}, \dot{\boldsymbol{r}}, t) = \frac{m\dot{\boldsymbol{r}}^2}{2} + q(\frac{1}{c}\boldsymbol{A} \cdot \dot{\boldsymbol{r}} - \phi)$$

---

[177] The Lagrangian is a function of a system's configuration and velocity. It can be expressed as the difference between kinetic and potential energy when conservative forces are at play, i.e., any work done in moving the particle is independent of the path taken. More generally, we can think of particular form of the Lagrangian as the one that is needed in order to obtain the correct dynamics of the system vis-à-vis the principle of least action. For an introduction to Lagrangian and Hamiltonian classical mechanics see Goldstein *et al*. (2000).

[178] E.g., Ballentine (1998, Ch. 11), Peebles (1992, Ch. 2.19), Gasiorowicz (2003 Ch. 16).

Where $\boldsymbol{r}$ is the particle's position in Cartesian coordinates ($\boldsymbol{r} = (r_x, r_y, r_z)$, $\dot{\boldsymbol{r}} = \frac{d}{dt}\boldsymbol{r}$ is the velocity of the particle so that $\dot{\boldsymbol{r}}^2 = \dot{\boldsymbol{r}} \cdot \dot{\boldsymbol{r}}$), and c is the constant speed of light in a vacuum. $\boldsymbol{A} = \boldsymbol{A}(\boldsymbol{r}, t)$ and $\phi = \phi(\boldsymbol{r}, t)$ are the electromagnetic vector and scalar potentials (respectively) such that the electric and magnetic fields can be expressed in the following manner:

$$\boldsymbol{B} = \nabla \times \boldsymbol{A}$$

$$\boldsymbol{E} = -\nabla\phi - \frac{\partial \boldsymbol{A}}{\partial t}$$

(where $\nabla = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z})$). The classical Hamiltonian H_{CM} of the system is defined as the Legendre transform of the Lagrangian:

$$H_{CM}(\boldsymbol{r}, \dot{\boldsymbol{r}}, t) = \boldsymbol{\Pi} \cdot \dot{\boldsymbol{r}} - L(\boldsymbol{r}, \dot{\boldsymbol{r}}, t)$$

Where the *canonical momentum* $\boldsymbol{\Pi}$ is defined as:

$$\boldsymbol{\Pi} = \frac{\partial L}{\partial \dot{\boldsymbol{r}}} = m\dot{\boldsymbol{r}} + \frac{q}{c}\boldsymbol{A}$$

(notice the dependence on $q\boldsymbol{A}$ in addition to the regular $m\dot{\boldsymbol{r}}$). Expressing $H(\boldsymbol{r}, \dot{\boldsymbol{r}}, t)$ in terms of the canonical momentum we get:[179]

$$H_{CM}(\boldsymbol{r}, \dot{\boldsymbol{r}}, t) = \frac{1}{2m}(\boldsymbol{\Pi} - \frac{q}{c}\boldsymbol{A})^2 + q\phi$$

In order to transition to standard non-relativistic quantum mechanics (QM) we must undertake the canonical (i.e., Hamiltonian) quantization procedure by which position and the canonical momentum are replaced with their (symmetric) operator counterparts.[180] In the coordinate representation (in the Schrödinger picture) this means that $\boldsymbol{\Pi}$ is replaced with the momentum operator $\boldsymbol{P} = -i\hbar\nabla$ ($i = \sqrt{-1}$, $\hbar = h/2\pi$, and h is Planck's constant.) The

[179] Where Hamilton's equations $\frac{d\boldsymbol{r}}{dt} = \frac{\partial H}{\partial \boldsymbol{\Pi}}; \frac{d\boldsymbol{\Pi}}{dt} = -\frac{\partial H}{\partial \boldsymbol{r}}$ lead to the familiar Lorentz force law $m\ddot{\boldsymbol{r}} = q(\boldsymbol{E} + \dot{\boldsymbol{r}} \times \boldsymbol{B})$, which can also be obtained by plugging the Lagrangian given here into the Euler-Lagrange equations $\frac{\partial L}{\partial \boldsymbol{r}} = \frac{d}{dt}\frac{\partial L}{\partial \dot{\boldsymbol{r}}}$.

[180] Operators acting on a separable Hilbert space (i.e., with a countable basis) $\mathcal{H}$.

electromagnetic vector $\boldsymbol{A} = \boldsymbol{A}(\boldsymbol{R}, t)$ and scalar $\phi = \phi(\boldsymbol{R}, t)$ potentials also become operators because they are functions of the position operator $\boldsymbol{R}$, where $\boldsymbol{P}$ and $\boldsymbol{R}$ satisfy the (Heisenberg form of the) canonical commutation relations: $[\boldsymbol{R}, \boldsymbol{P}] \equiv \boldsymbol{RP} - \boldsymbol{PR} = i\hbar$.[181] The Hamiltonian operator now becomes:

$$H = \frac{1}{2m}(\mathbf{P} - \frac{q}{c}\boldsymbol{A})^2 + q\phi$$

(where I have hidden the dependence on position and time for convenience). If we then set the scalar potential to zero, we arrive at the form of the Hamiltonians H^I (which is not self-adjoint nor essentially self-adjoint on $C^\infty(\mathbb{R}^3 \setminus S_{in})$, i.e., it does not have a unique self-adjoint extension on the domain of smooth compactly supported functions on the configuration space that arises when one removes the infinite cylinder from $\mathbb{R}^3$) and H^I_{AB} (which is the particular self-adjoint extension of H^I used by Aharonov and Bohm (1959) and picked out by Dirichlet boundary conditions).

[181] And $[\boldsymbol{R}, \boldsymbol{R}] = [\boldsymbol{P}, \boldsymbol{P}] = 0$. As opposed to the story told here, where one starts with the action principle and the Lagrangian of some system, many physicists take the commutation relations as the fundamental structure upon which one can build a quantum theory.

APPENDIX B. THE RECEIVED TEXTBOOK ACCOUNT OF THE AB EFFECT

We want to first consider a double-slit experimental situation, and subsequently extend the discussion to an idealized AB effect set-up wherein we add a solenoid, such as in Figure 6.2. In a double-slit experiment, charged particles (such as electrons) are emitted from a source and directed at a double-slit. The particle beam is split up as it traverses through the slits, and then recombines at a detector screen. We can consider two (overlapping) regions, 1 and 2, one relating to the beam traversing path 1, represented by the state function Ψ_1, and one corresponding to path 2 and Ψ_2 (as in Figure 6.2). Generally, the interference pattern in a double-slit experiment (without a solenoid) arises because the of a relative phase difference between the two beams. For instance, if we represent the beams by a three-dimensional plane waves, then they will take the following form:

$$\Psi(\boldsymbol{r}, t) = Ae^{i\theta}$$

Where A is the amplitude of the wave, θ is its phase, and Ψ is related to the intensity I of the wave via:

$$I = |\Psi|^2$$

Let I_1 be the intensity at the detector screen when only the slit related to path 1 is open, I_2 will similarly be the intensity related to the slit of path 2, and the I will be the total intensity collected at the screen. We then have (for waves) the following relation for the total intensity (Zettili 2009,23):

$$I = |\Psi_1 + \Psi_2|^2 = |\Psi_1|^2 + |\Psi_2|^2 + (\Psi_1^*\Psi_2 + \Psi_1\Psi_2^*) =$$

$$= I_1 + I_2 + 2\sqrt{I_1 I_2}\cos\delta$$

Where * denotes the complex conjugate, and δ is the phase difference between Ψ_1 and Ψ_2 responsible for the interference pattern. The condition for constructive interference is $\delta = 2\pi n$ (for integer n), and δ is related to the geometry of the experimental apparatus as follows:

$$\delta \approx \frac{2\pi x d}{\lambda l}$$

x is the distance from the axis of symmetry (the middle of the screen) where an interference maxima appears, l is the distance to the screen, λ is the de Broglie wavelength of the electrons in the beam, d is the slit separation, and we are considering the limit in which $x \ll l$ (as is the case in experiment). Thus, an interference maxima appears at $x \approx n\lambda l/d$. However, if we now shift to an AB effect experimental set-up, where a solenoid is added, the phase difference is affected by the total magnetic flux, so that the maxima is shifted by an amount:

$$\Delta x = \frac{l\lambda}{2\pi d}\Delta\delta$$

where $\Delta\delta$ is the novel phase difference due to the presence of the solenoid. This is the (idealized) physical manifestation of the AB effect.

Returning to the Hamiltonian H (from Appendix A) of a charged particle in a magnetic field, the usual account of $\Delta\delta$ and its relation to the total magnetic flux runs as follows. Let $\Psi(\boldsymbol{r}, t)$ be the solution to the time-dependent Schrödinger equation $H\Psi = i\hbar\frac{\partial}{\partial t}\Psi$. In the same way that the electromagnetic fields are gauge invariant, it is known that Schrödinger equation

$$\frac{1}{2m}(\mathbf{P} - \frac{q}{c}\boldsymbol{A})^2\Psi + q\phi\Psi = i\hbar\frac{\partial}{\partial t}\Psi$$

is unchanged by the following gauge transformation:

$$\boldsymbol{A} \rightarrow \boldsymbol{A}' = \boldsymbol{A} + \boldsymbol{\nabla}\Lambda$$

$$\phi \rightarrow \phi' = \phi - \frac{\partial \Lambda}{\partial t}$$

$$\Psi \rightarrow \Psi' = \Psi e^{i\frac{q}{c\hbar}\Lambda}$$

where $\Lambda = \Lambda(\mathbf{r}, \mathrm{t})$ is an arbitrary scalar function.

Consider a current-carrying, closely wound, cylindrical solenoid with radius r_0, that is infinitely long and absolutely impenetrable (see Figure 6.2). Infinite length and impenetrability imply that the constant magnetic field $\boldsymbol{B}$ produced by the current is wholly confined to a region $\mathcal{S}_{in}$ inside the solenoid, so that particles in the region $\mathcal{S}_{out}$ outside the solenoid have no contact with $\boldsymbol{B}$. In particular, the assumption that the solenoid is infinite guarantees that there is no spillage of magnetic field into the region $\mathcal{S}_{out}$ outside the solenoid. Allegedly, the consequence of the impenetrability assumptions is that state functions vanish $\Psi = 0$ at the solenoid boundary. However, as discussed in Chapter 6, the impenetrability assumption only guarantees that the electron probability current vanishes at the solenoid boundary $j \coloneqq -i(\Psi^* \nabla \Psi - \Psi \nabla \Psi^*)$, but this can be implemented using *different* boundary conditions including the Dirichlet boundary conditions ($\Psi = 0$) that arise in standard textbook accounts, Neuman boundary conditions ($\nabla \Psi = 0$), or Robin boundary conditions ($\nabla \Psi = r\Psi, r \in \mathbb{R}$) (de Oliviera and Pereira 2010). Setting such details aside, by picking Dirichlet boundary conditions, and given physical situation at hand is one that consists of both an electron beam and a magnetic field, $H_{AB}^{I} = \frac{1}{2m}(\mathbf{P} - \frac{q}{c}\boldsymbol{A})^2$ becomes the appropriate Hamiltonian to use in representing the region outside the solenoid.

The solenoid is centered at the origin and in the z-axis direction. For concreteness, notice that the following electromagnetic vector potential (in cylindrical coordinates (r, θ, z)) can give rise to a magnetic field inside the solenoid $\boldsymbol{B} = (0,0,B)$, which vanishes in the region outside the solenoid:

$$A_z = A_r = 0; A_\theta = \frac{\Phi}{2\pi r}\ (r \geq r_0 > 0)$$

$$A_z = A_r = 0; A_\theta = \frac{\Phi}{2\pi r_0^2} r\ (r_0 \geq r > 0)$$

Where $\Phi = \iint_S \boldsymbol{B} \cdot d\boldsymbol{r^2} = \pi r_0^2 B$ is the magnetic flux.

Recall that we are considering two regions 1 and 2, with state functions Ψ_1 and Ψ_2 representing the beams through the two slits. We can use the above gauge transformations to express the state functions as $\Psi_1 = \Psi_1{}' e^{i\frac{q}{c\hbar}\Lambda_1}$ and $\Psi_2 = \Psi_2{}' e^{i\frac{q}{c\hbar}\Lambda_2}$, where $\Psi_1{}'$ and $\Psi_2{}'$ are the zero-potential solutions ($\boldsymbol{A} = 0$) to the time-dependent Schrödinger equation with H_{AB}^I. We choose $\Lambda_1 = \int_1 \boldsymbol{A} \cdot d\boldsymbol{r}$ and $\Lambda_2 = \int_2 \boldsymbol{A} \cdot d\boldsymbol{r}$, where the line integrals are evaluated about paths 1 and 2, respectively. The state function at the detector screen will be $\Psi = \Psi_1{}' e^{i\frac{q}{c\hbar}\Lambda_1} + \Psi_2{}' e^{i\frac{q}{c\hbar}\Lambda_2}$. The interference between the two beams will manifest as a pattern on the screen that depends on the relative phase between to two paths $e^{i\frac{q}{c\hbar}(\Lambda_1 - \Lambda_2)}$, but this is equivalent to evaluating the integral around a closed curve C (half of which corresponds to taking path 1, and half of which corresponds to taking the reverse of path 2): $\Lambda_1 - \Lambda_2 = \oint_C \boldsymbol{A}(\boldsymbol{r}) \cdot d\boldsymbol{r}$. We can use Stoke's theorem to get the following:

$$\oint_{C=\partial S} \boldsymbol{A}(\boldsymbol{r}) \cdot d\boldsymbol{r} = \iint_S (\nabla \times \boldsymbol{A}) d\boldsymbol{r^2} = \iint_S \boldsymbol{B} \cdot d\boldsymbol{r^2} = \Phi$$

where the double integral is evaluated with respect to the surface area S enclosed by curve C. Hence, $\Delta\delta \equiv \frac{q}{c\hbar}(\Lambda_1 - \Lambda_2) = \frac{q}{c\hbar}\Phi$ and total shift in the maxima is:

$$\Delta x = \frac{l\lambda q}{2\pi d c\hbar}\Phi$$

APPENDIX C. FIBER BUNDLE FORMULATION OF THE AB EFFECT

For visualization purposes it would help to consider a (finite) cylinder and also a Möbius strip as two examples of fiber bundle structure that I will refer to (see Figure 8.3) A (differentiable) fiber bundle (or bundle for short) is a five-tuple (E, π, M, F, G). E, M and F are differentiable manifolds (and hence topological spaces) known as the *total space*, the *base space*, and the *typical fiber* (or fiber for short), respectively. In the context of our examples of the cylinder and Möbius strip, the base space is a (unit) circle $M = S^1$, the typical fiber is a line segment (say) $F = \langle 0,1 \rangle$, and the total space is the cylinder or the Mobius strip themselves (where the equality symbol "$=$" between spaces represents a homeomophism). In the case of the cylinder $E = M \times F$ (where "$\times$" denotes the Cartesian product), we say that the bundle is *trivial*. In contrast, the Möbius strip has the same structure at the cylinder locally but globally it also has a twist so that the bundle is *non-trivial*. π is a *projection* map from the total space to the base space, $\pi: E \rightarrow M$, that associates points $\tilde{q} \in M$ in the base space with points $q \in E$ in the total space E constituting the *fiber above* $\tilde{q}$, $F_{\tilde{q}}$ ($F_{\tilde{q}} \subseteq E$), where $F_{\tilde{q}}$ has the structure of the typical fiber. In other words, all the points in the fiber above $\tilde{q}$ in the total space get mapped to the point $\tilde{q} \in M$ in the base space via π. G is a group of homeomorphisms of the fiber F (mapping every $F_{\tilde{q}}$ onto itself) , known as the *structure group* of the fiber bundle or the *gauge group* (of the theory being formulated in terms of fiber bundles). A *cross-section* (or section for short) σ of E is a continuous map from subsets of the base space to the total space $\sigma: U \subseteq M \rightarrow E$ such that $\pi \circ \sigma = I$ is the identity map (on U). We say that a section is *global* if $U = M$, otherwise it is *local*. A *vector bundle* (associated with E) is a fiber bundle where the typical fiber constitutes a vector space. For instance, $F = \mathbb{R}$ (with $M = S^1$) corresponds to an infinite cylinder. A *principal fiber bundle* (or principal bundle for short) (associated with E) $P(M, G)$ is a vector bundle in which the typical fiber and the structure group are the same: $F = G$. A basic result is that the total space E of the fiber bundle and its corresponding principal bundle $P(M, G)$ are trivial if and only if $P(M, G)$ has a global section.

The fiber bundle formulation of classical electromagnetism takes the base space to be the space ($M = \mathbb{R}^3$; where $\mathbb{R}$ is the set of real numbers) or spacetime manifold ($M = \mathbb{R}^4$), and the group structure to be the multiplicative group of complex numbers of modulus 1: $G = U(1)$. The corresponding principal bundle is $P(\mathbb{R}^4, U(1))$, or P for short. The *connection* ω is a Lie-algebra-valued one-form on P that basically pairs all smooth curves $[\tilde{C}]$ through $\tilde{q} \in M$ with a corresponding class of smooth curves $[C]$ in E, known as *horizontal lifts*. If $\tilde{C}$ is a closed curve based at $\tilde{q} \in M$, the connection defines a horizontal lift to some curve C in E that begins at q and ends at q' ($q, q' \in F_{\tilde{q}} \subseteq E$). The concept of curvature Ω, which is a Lie-algebra-valued two-form on P, is introduced as follows. If $q = q'$ we say the curvature is flat, but if $q \neq q'$ we say bundle is curved. The bundle section maps Ω to another two-form iF' (on a subset of M) where F' can be identified with the magnetic field $\boldsymbol{B}$ (in the purely magnetic case). (I use F' so as not confuse with the typical fiber F). Similarly, the connection ω is mapped by the bundle section to another one-form iA (on an open set of M) where A can be identified with the vector potential $\boldsymbol{A}$.

Recall, in the context of the AB effect, observable effects such as interference shifts are determined by the phase factor attached to the wave function:

$$\exp(\frac{iq}{\hbar}\oint_{\tilde{C}} \boldsymbol{A} \cdot d\boldsymbol{r})$$

The wave function is represented by a global section of the vector bundle $(E, \pi_E, \mathbb{R}^4, \mathbb{C}, U(1))$ associated to the principal fiber $P(\mathbb{R}^4, U(1))$ (where π_E is a projection map from the associate vector bundle to the base space such that each point $p \in E$ is an equivalence class of points $[(q, c)]$ where $q \in P$, $c \in \mathbb{C}$, and p is mapped onto the same point m that q is mapped via π; $\mathbb{C}$ is the set of complex numbers). The phase factor, then, arises naturally in the fiber bundle formalism as the *holonomy* of the curve $\tilde{C}$ in the following manner. The connection essentially maps the fiber above m onto itself via a horizontal lift. If $\tilde{C}$ starts and ends at m, then the curve is lifted to the bundle with start and end points $q, q' \in F_{\tilde{q}} \subseteq E$ (as explained above). It turns out that for classical electromagnetism the map is induced by the action of (a matrix representation of) an element—called the holonomy—of the structure group $U(1)$, which is independent of $\tilde{C}$'s starting point and of the section.

APPENDIX D. VISUALIZING FRACTIONAL STATISTICS ON THE CONFIGURATION SPACE APPROACH

The fundamental group of the configuration space of the simplest scenario of two particles $N = 2$ in the $d = 2$ and $d = 3$ cases is as follows:

$$\pi_1 \left(\frac{\mathbb{R}^2 - \Delta}{S_2} \right) = \pi_1 \left(RP_1 \right) = Z \text{ for } d = 2$$

$$\pi_1 \left(\frac{\mathbb{R}^3 - \Delta}{S_2} \right) = \pi_1 \left(RP_2 \right) = Z_2 \text{ for } d = 3$$

Where Z is the cyclic group of order one, i.e., the infinite group of integers under addition. Z_2 is the cyclic group of order two, i.e., it is the multiplicative group of, say, 1 and -1. RP_1 and RP_2 are the real projective one- and two-dimensional spaces, respectively.

Pictorially, for the $d = 3$ case the configuration space reduces to the real projective space in two dimensions RP_2. This can be visualized as the surface of a three-dimensional sphere with diametrically opposite points identified (see Figure D.1). Consider three scenarios, corresponding to three paths A, B, and C in configuration space including no exchange (Figure D.1a), exchange (Figure D.1b), and a double exchange (Figure D.1c), respectively.

Concentrating on the no exchange case (Figure D.1a). We trace a path A in configuration space in which the two particles move and return to their original positions. Path A is a loop in configuration space, with the same fixed start and end points, which can be shrunk to a point. This corresponds to a trivial homotopy class in which the phase factor is trivial.

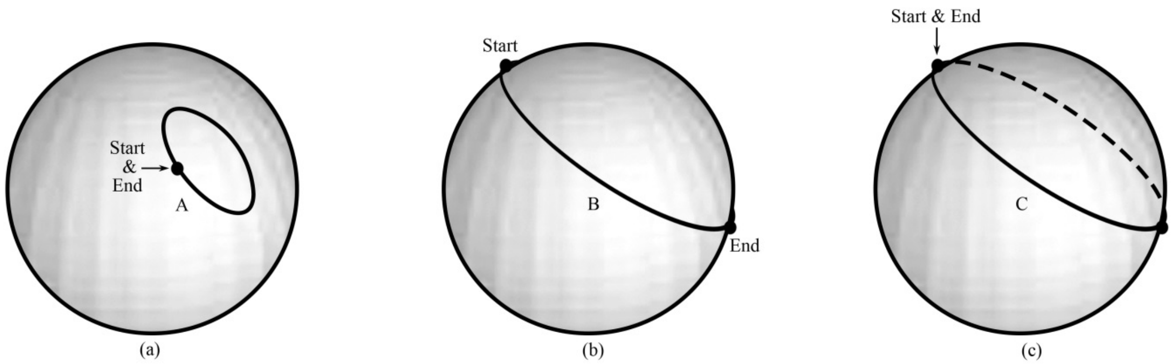


Figure D.1: The real projective space in two dimensions RP_2, represented by a sphere with diametrically opposite points identified. Cases (a), (b), and (c), correspond to no exchange, exchange, and double exchange, respectively.

Moving onto the exchange case (Figure D.1b), we start at one end of the configuration space and trace a path B to its diametrically opposite point. This represents an exchange or permutation between the two particles. Notice that since diametrically opposite points are identified (because the particle are identical), this path is actually a closed loop in configuration space. However, since the start and end points of Figure D.1b are fixed, the loop cannot be shrunk to point. This corresponds to a non-trivial homotopy class with a non-trivial phase factor.

The double exchange (Figure D.1c) case includes tracing a path C in configuration space similar to that of B, but then tracing around the sphere back to the original starting point. Path C is a closed loop in configuration space that can be shrunk to a point, and so it is in the same homotopy class of path A with a corresponding trivial phase factor. Equivalently, we may visualize the paths A, B, C on a hemisphere with opposite points on the equator identified as in

Figure D.2, where paths A and C can be continuously deformed to a point but path B cannot because of the diametrically opposed fixed start and end point on the equator.

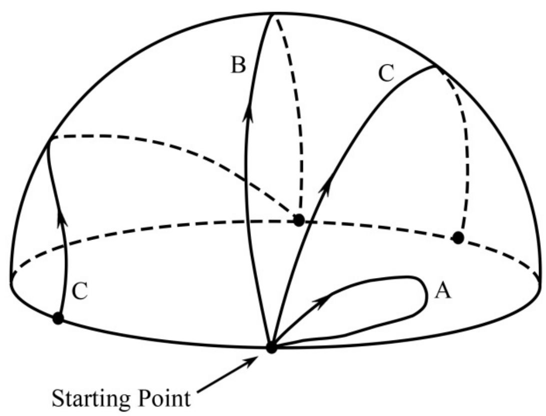


Figure D.2: The real projective space in two dimensions RP_2, represented by the northern hemisphere with opposite point on the equator identified.

On the other hand, in the context of the $d = 2$ case, we are dealing with the real projective space in one dimension RP_1. We can visualize this configuration space as a circle with diametrically opposite points identified (see Figure D.3). Again, consider three paths A, B, and C in configuration space that correspond to no exchange (Figure D.3a), exchange (Figure D.3b), and a double exchange (Figure D.3c), respectively. Path A traces a closed loop in configuration space (where the particles move but then return to their original positions with no exchange) which can be continuously shrunk to a loop and has a corresponding trivial phase factor (as in the $d = 3$ case of Figure D.1a). Next, we trace a path B across half the circumference of the circle. Since diametrically opposed points are identified, this represents a particle exchange (Figure D.3b). Path B traces a closed loop in configuration space that cannot be continuously shrunk to a loop and has a corresponding non-trivial phase factor (as in the $d = 3$ case of Figure D.1b).

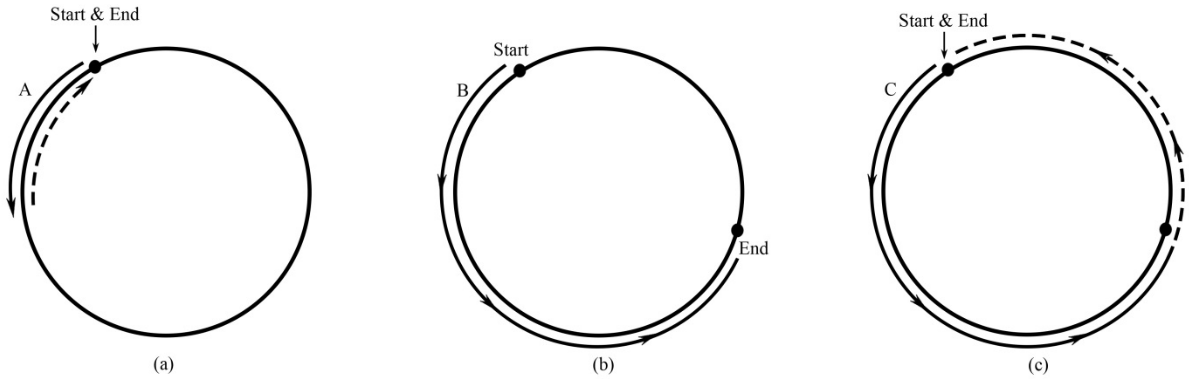


Figure D.3: The real projective space in one dimension RP_1, represented by a circle with diametrically opposite points identified. Cases (a), (b), and (c), correspond to no exchange, exchange, and double exchange, respectively.

The main difference between the $d = 3$ and $d = 2$ cases arises when we consider path C (Figure D.3c), in which the particles are permuted twice, represented by traversing the entire circular configuration space. Path C is a closed loop in configuration space but, unlike the $d = 3$ case, it cannot be shrunk to a loop because the circle itself (so to say) acts as an obstructive barrier. Moreover, path C cannot even be continuously deformed to overlap with path B. This means that, not only is the phase factor corresponding to the two paths non-trivial, but each path has a different phase factor for each path belongs to a different homotopy class. In fact, for every traversal (in configuration space) of half a circle, we get a closed loop that is in its own homotopy class.[182] In other words, by transitioning from three dimensions to two dimensions, we

[182] If we symbolize this by π_1 ($Path$) we get that π_1 $(Path \;\; A) = 0$ for the trivial homotopy class, but the rest of the paths will be elements of non-trivial homotopy classes: π_1 $(Path\ B) = 1, \pi_1$ $(\ Path\ C) = 2, \ldots$ and so on, so that we

have transitioned from a doubly connected space to an infinitely connected space, and it is this change in topology that allows for fractional statistics.

---

generate all of the integers Z. Negative integers corresponding to traversal of the circular configuration space in the opposite direction.

APPENDIX E: A GEOMETRIC APPROACH TO FRACTIONAL STATISTICS

We follow Berry's (1984) original paper on the non-dynamical phase factor accompanying cyclic evolutions of quantum systems. To begin, consider some system, such as a spinless electrically charged particle in a box, with a corresponding Hamiltonian $H\big(\boldsymbol{R}(t)\big)$ that depends on a set of parameters $\boldsymbol{R} = (X, Y, \dots)$ and can be altered over time by varying said parameters. We can view the alteration of $H\big(\boldsymbol{R}(t)\big)$ as a path in parameter space. If the system starts at some time $t = 0$, and is gradually changed over time t so that the parameter values are returned to their original values $\boldsymbol{R}(0) = \boldsymbol{R}(t)$ then this maps out a closed curve C in parameter space. According to the *adiabatic theorem*, if the system was originally (at time $t = 0$) in the nth eigenstate $\psi_n(\boldsymbol{R}(0))$ of $H\big(\boldsymbol{R}(0)\big)$, if $H\big(\boldsymbol{R}(t)\big)$ is non-degenerate, and if the excursion in parameter space is sufficiently slow, then the system will transition (under Schrödinger evolution) into the nth eigenstate $\psi_n(\boldsymbol{R}(t))$ of $H\big(\boldsymbol{R}(t)\big)$ (with some added overall phase factor).[183]

The general state of the system $\Psi(t)$ evolves according to the time-dependent Schrödinger equation, and at *any instant* t the eigenstates of the time-independent Schrödinger equation form a natural basis satisfying:

$$H\big(\boldsymbol{R}(t)\big)\psi_n(\boldsymbol{R}(t)) = E_n(\boldsymbol{R}(t))\psi_n(\boldsymbol{R}(t))$$

---

[183] The adiabatic theorem is originally due to Born and Fock (1928). See, for instance, Bransden and Joachaim (2000, Section 9.4) for a proof.

According to the *adiabatic approximation* then, the general state of the system $\Psi(t)$ at some time t can be expressed as follows:[184]

$$\Psi(t) = \psi_n(\boldsymbol{R}(t))e^{i\theta}$$

Where the exchange phase θ has two components $\theta = \theta_D + \theta_G$ such that:

$$\theta_D = -\frac{1}{\hbar}\int_0^t E_n(t)dt \text{ and } \theta_G = i\int_0^t \left\langle \psi_n(t) \middle| \frac{\partial}{\partial t}\psi_n(t) \right\rangle dt \tag{E.1}$$

θ_D corresponds to the usual *dynamical phase* (accompanying the Schrödinger evolution of any stationary state) and θ_G is called the *geometric phase* or *Berry's phase* (where I have used Dirac's bra-ket notation and hid the parameter dependence for convenience). It can be expressed more generally as a quantity dependent on both the closed curve C in parameter space and the parameters $\boldsymbol{R} = (X, Y, \dots)$:

$$\theta_G(C) = i\oint_C \langle \psi_n(\boldsymbol{R}) | \nabla_R \psi_n(\boldsymbol{R}) \rangle \cdot d\boldsymbol{R} \tag{E.2}$$

Where ∇_R is the gradient with respect to the parameters $\boldsymbol{R} = (X, Y, \dots)$ (and assuming that $\boldsymbol{R}(0) = R(\boldsymbol{t})$ so that C forms a closed curve). There are similar results for degenerate systems (Wilczek and Zee 1984) and for a cyclic evolution that is not necessarily adiabatic (Aharonov and Anandan 1987).

Next, my goal in the rest of this appendix is solely to repeat some of the steps taken by Arovas, Schrieffer, and Wilczek (1984) to derive fractional statistics in order to emphasize the disconnect between this geometric approach and the topological (and pathological) approach discussed in Section 7.3. Many steps will be skipped, and I refer the reader interested in a more details to explicit calculations made by by Arovas (1989) in Wilczek and Shapere (1989, 284-322) and Laughlin (1990) in Wilczek (1990, 262-303).

[184] See, for instance, Griffiths (2005, 373) for the adiabatic approximation. Basically, we drop terms that depend on the time or parameter derivative of $H(\boldsymbol{R}(t))$ for, by assumption, the change is minute.

Following Laughlin (1983a, 1983b) and Arovas, Schrieffer, and Wilczek's (1984) closely, let us consider a FQHE system with filling factor $\nu = \frac{1}{m}$ where m is an odd integer, and the applied strong magnetic field $\boldsymbol{B}$ is in the z-*axis* direction corresponding to magnetic flux Φ. In such a situation, the Hamiltonian governing the system is:[185]

$$H_{FQHE}^{I} = \sum_{j} \frac{\left(\boldsymbol{p_j} - q\boldsymbol{A_j}\right)^2}{2m_e} + \boldsymbol{V}(z_j) + \sum_{j>k} \frac{e^2}{\left|z_j - z_k\right|}$$

Recall, $z_j = x_j + iy_i$ are in units of magnetic length $l_B = \sqrt{\hbar\, c/eB}$, which have been set to equal one, e is the charge of the electron, and j and k run over N particles. The $\frac{\left(\boldsymbol{p_j} - q\boldsymbol{A_j}\right)^2}{2m_e}$ term signifies the kinetic energy of charged particles in a magnetic field, $\boldsymbol{V}(z_j)$ is average background potential, and $\frac{e^2}{|z_j - z_k|}$ is the Coulomb interaction between particles. Laughlin's (1983a, 1983b) celebrated wavefunction for the ground state of H_{FQHE}^{I} is:

$$\Psi_m = \prod_{j<k}^{N} (z_j - z_k)^m e^{\left(-\frac{1}{4}\sum_l^N |z_l|^2\right)}$$

The state function of two excited states (quasiholes) a and b located at positions z_a and z_b, respectively, is represented by

$$\Psi_m^{z_a z_b} = N_{ab} \prod_i (z_i - z_a)(z_i - z_b)\Psi_m \qquad \text{(E.3)}$$

where N_{ab} is a normalizing factor.

We can determine the quantum statistics associated with exchanging quasiholes a and b by calculating the geometric phase associated with carrying quasihole a adiabatically around a closed loop C, thereby adding time dependence to $z_a = z_a(t)$, and identifying the geometric

---

[185] I use the H_{FQHE}^{I} notation to emphasize that this is the idealized (I) Hamiltonian corresponding to FQHE systems.

phase with the exchange phase. The geometric phase θ_G can be calculated by plugging Equation. E.3 into Equation E.1 or Equation E.2 as follows:

$$\frac{d\theta_G}{dt} = i\left\langle \Psi_m^{z_a z_b}(z_a(t), z_b) \middle| \frac{d}{dt} \middle| \Psi_m^{z_a z_b}(z_a(t), z_b) \right\rangle \quad \text{(E.4)}$$

Denoting the mean number of electrons inside loop C with $\langle n_e \rangle_C$, it turns out that solving Equation E.4 leads to the following expression for the geometric phase $\theta_G = -2\pi \langle n_e \rangle_C$.[186] If quasihole b is outside the loop then $\langle n_e \rangle_C$ is equal to $\frac{\nu\Phi}{\Phi_0}$, where ν is the filling factor, Φ is the magnetic flux corresponding to the strong magnetic field applied in FQHE systems, and the constant $\Phi_0 = \frac{hc}{e}$ is the "flux quanta," so that $\theta_G = -2\pi \frac{\nu\Phi}{\Phi_0}$. However, if quasihole b is inside the loop then there is a deficit in mean number of electrons by an amount $-\nu$ so that $\theta_G = -2\pi \frac{\nu\Phi}{\Phi_0} + 2\pi\nu$. The relative difference in geometric phase between the two scenarios is $\Delta\theta_G = 2\pi\nu$.

In other words, when quasihole a encircles quasihole b, the new doubly permuted wavefunction $\psi_m'^{z_a z_b}$ gains an extra geometric phase $\Delta\theta_G = 2\pi\nu$:

$$\Psi_m'^{z_a z_b} = e^{i2\pi\nu} \Psi_m^{z_a z_b}$$

But recall from Section 7.1 that double permutation leads to a general overall phase factor with an exchange phase θ:

$$\Psi_m'^{z_a z_b} = e^{i2\theta} \Psi_m^{z_a z_b} = e^{i2\pi\nu} \Psi_m^{z_a z_b} = e^{i2\pi\alpha} \Psi_m^{z_a z_b}$$

Where we have introduced the "statistical parameter" defined as $\alpha = \frac{\theta}{\pi}$. We see that $\alpha = \nu$ and recalling that $\nu = \frac{1}{m}$ where m is an odd integer, it follows that $\theta = \frac{\pi}{m}$. For the m=1

---

[186] See Shapere and Wilczek (1989, 307-309) and Wilczek (1990, 300-301) for an explicit calculation.

case, $\theta = \pi$ corresponding to Fermi-Dirac statistics. But for other values of m, θ corresponds to anyonic statistics.

BIBLIOGRAPHY

Aharonov, Y. and Bohm, D. 1959. “Significance of Electromagnetic Potentials In the Quantum Theory.” *Physical Review*. 115: 485-491.

Ando, T., Fowler, A. B., Stern, F. 1982. “Electronic Properties of Two-dimensional Systems.” *Review of Modern Physics*. 54: 437-672.

Arovas, D. P. 1989. *Topics in Fractional Statistics*. In: Shapere, A., Wilczek, F. (eds.) 1989. *Geometric Phases in Physics*. Singapore: World Scientific. 284-322.

Arovas, D. P., Schrieffer, J. R., Wilczek, F. 1984. “Fractional Statistics and the Quantum Hall Effect.” *Physical Review Letters* 53: 722-723.

Artin, E. 1947. “Theory of Braids.” *Annals of Mathematics*. 48(1): 101-126.

Azzouni, J. 2012. “Taking the Easy Road Out of Dodge.” *Mind*. 121: 951–66.

Babiker, M., Loudon, R. 1984. “Gauge Invariance of the Aharonov-Bohm Effect.” *Journal of Physics A: Mathematical and General* 17: 2973–2982.

Baker, A. 2005. “Are there Genuine Mathematical Explanations of Physical Phenomena?” *Mind*. 114: 223–38.

Baker, A. 2009. “Mathematical Explanation in Science.” *British Journal for the Philosophy of Science*. 60: 611–633.

Baker, A. 2012. “Science-Driven Mathematical Explanation.” *Mind*. 121: 243–267.

Balaguer, M. 1998. *Platonism and Anti-Platonism in Mathematics*. New York: Oxford University Press.

Ballentine, L. E. 1998. *Quantum Mechanics: A Modern Development*. Singapore: World Scientific.

Ballesteros, M., R. Weder. 2009a. “The Aharonov–Bohm Effect and Tonomura et al. Experiments: Rigorous Results.” *Journal of Mathematical Physics* 50:122108.

Ballesteros, M., R. Weder. 2009b. "High-velocity estimates for the scattering operator and Aharonov–Bohm effect in three dimensions." *Communications in Mathematical Physics* 285: 345–398.

Ballesteros, M., R. Weder. 2011. "Aharonov–Bohm effect and high-velocity estimates of solutions to the Schrodinger equation." *Communications in Mathematical Physics* 303(1): 175-211.

Bangu, S. 2009. "Understanding Thermodynamic Singularities: Phase Transitions, Date and Phenomena." *Philosophy of Science* 76:488-505.

Bartels, A. 2006. "Defending the Structural Conception of Representation" *Theoria.* 50: 7-19.

Batterman, R. W. 1992. "Explanatory Instability." *Nous*, 26(3):325-348.

Batterman, R. W. 2000. "Multiple Realizability and Universality." *British Journal for the Science.* 51(1): 115-145.

Batterman, R. W. 2002. *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence*. New York: Oxford University Press.

Batterman, R. W. 2003. "Falling cats, parallel parking, and polarized light." *Studies in History and Philosophy of Modern Physics*. **34:** 527–557.

Batterman, R. W. 2005. "Critical Phenomena and Breaking Drops: Infinite Idealizations in Physics." *Studies in History and Philosophy of Modern Physics.* 36: 225-244.

Batterman, R. W. 2009. "Idealization and Modeling." *Synthese.* 169(3): 427-446.

Batterman, R. W. 2010. "On the Explanatory Role of Mathematics in Empirical Science." *British Journal for the Science*. 61(1): 1-25.

Batterman, R. W. 2013. "The Tyranny of Scales." In R. W. Batterman, ed., *The Oxford Handbook of Philosophy of Physics*. Oxford University Press. 255-286.

Batterman, R. and Rice, C. 2014. "Minimal Model Explanations." *Philosophy of Science.* 81(3): 349-376.

Belot, G. 1998. "Understanding electromagnetism." *British Journal for Philosophy of Science* 49(4): 531–555.

Berry, M. V. 1984. "Quantal Phase Factors Accompanying Adiabatic Changes." *Proceedings of the Royal Society A.* A392: 45-57.

Berry, M. V. 1986. “The Aharonov-Bohm Effect is Real Physics Not Ideal Physics.” In *Fundamental Aspects of Quantum Theory*, eds. Gorini, V., A. Frigerio, 144: 319–320. New York: Plenum.

Birman, J. S., Brendle, T. E. 2005. “Braids: A Survey.” Available via arXiv:math/0409205

Bocchieri, P., Loinger, A. 1978. “Nonexistence of the Aharonov-Bohm Effect.” *Nuovo Cimento A* 47: 4750-4482.

Bocchieri, P., Loinger, A., Siragusa, G. 1979. “Nonexistence of the Aharonov-Bohm Effect. II Discussion of the Experiments.” *Nuovo Cimento A* 51: 1.

Bocchieri, P., Loinger, A., Siragusa, G. 1980. “The Role of the Electromagnetic Potentials in Quantum Mechanics. The Marton Experiment..” *Nuovo Cimento A* 56: 55.

Bohm, D., Hiley, B. J. 1979. “On the Aharonov-Bohm Effect.” *Nuovo Cimento A* 52: 295.

Bokulich, A. 2008. *Re-examining the Quantum-Classical Relation: Beyond Reductionism and Pluralism*. Cambridge University Press. (2008)

Bolinska, A. 2013. “Epistemic Representation, Informativeness and the Aim of Faithful Representation.” *Synthese.* (190): 219–234.

Borrmann, P., Mülken, O., and Harting, J. 2000. “Classification of Phase Transitions in Small Systems.” *Physical Review Letters*. 84(16), 3511.

Boyd, R. N. 1983. “On the Current Status of the Issue of Scientific Realism.” *Erkenntnis* 19:45–90.

Boyd, R., 1984. “The Current Status of Scientific Realism,” in J. Leplin (ed.), *Scientific Realism*, Berkeley CA: University of California Press, pp. 41–82.

Bueno, O. 1997. “Empirical Adequacy: A Partial Structure Approach” *Studies in History and Philosophy of Science* 28:59-74.

Bueno, O. 2012. “An Easy Road to Nominalism.” *Mind*. 121: 967–82.

Bueno, O., Steven, F., Ladyman, J. 2002. “On Representing the Relationship Between the Mathematical and the Empirical.” *Philosophy of Science*, 69:497-518.

Brandom, R. 1994. *Making it explicit*. Cambridge, MA: Harvard University Press.

Bransden, B. H., Joachaim, J. C. 2000. *Quantum Mechanics*. Prentice Hall.

Brodbeck, M. 1968. “Theory Construction.” In *Readings in the Philosophy of Science*. New York: The MacMilian Company.

Brzezinski, J. and Nowak, L. (eds.) 1992. *Idealization III: Approximation and Truth*. Amsterdam: Rodopi.

Burton, D. 2011. *The History of Mathematics: An Introduction*. McGraw-Hill Higher Education.

Butterfield, J. 2011. “Less is Different: Emergence and Reduction Reconciled.” *Foundations of Physics* 41(6): 1065-1135.

Butterfield, J. N. 2006. “On Symmetry and Conserved Quantities in Classical Mechanics.” In W. Demopoulos and I. Pitowksy (eds.) *Physical Theory and Its Interpretation: Essays in Honor of Jeffrey Bub*. Springer.

Callender, C. 2001. “Taking Thermodynamics Too Seriously.” *Studies in History and Philosophy of Science Part B: Modern Physics* 32(4), 539–53.

Callender, C., and Cohen, J. 2006. “There Is No Special Problem About Scientific Representation.” *Theoria*. 50: 67-85.

Camino, F. E., Zhou, W., Goldman, V. J. 2005. “Realization of a Laughlin Quasiparticle Interferometer: Observation of fractional statistics.” *Physical Review B***.** 72: 075342.

Canright, G. S., Johnson, M. D. 1994. “Fractional Statistics: α to β.” *Journal of Physics A: Mathematical and General*. 27(11): 3579-3598.

Caprez, A., Barwick, B.B., Batelaan, H. 2007. “Macroscopic test of the Aharonov–Bohm effect.” *Physical Review Letters*. 99: 210401.

Cartwright, N. 1983. *How the Laws of Physics Lie*. Oxford: Clarendon Press.

Cartwright, N. 1989. *Nature's Capacities and Their Measurement.* Oxford: Clarendon Press.

Cartwright, N. 1999. *The Dappled World: A Study of the Boundaries of Science*. New York: Cambridge University Press.

Cartwright, N. 2000. “Against the Completability of Science.” In M. W. F. Stone and J. Wolff (eds.) *The Proper Ambition of Science*. London: Routlege.

Chakraborty, T., Pietilinen, P. 1995. *The Quatnum Hall Effects*., Berlin: Springer.

Chakravartty, A. 2001. “The Semantic or Model-Theoretic View of Theories and Scientific Realism.” *Synthese*, 127(3).

Chakravartty, A. 2010. “Informational Versus Functional Theories of Scientific Representation. *Synthese.* 172(2), 197-213.

Chakravartty, A. 2011. “Scientific Realism”. In *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta (ed.) URL= http://plato.stanford.edu/entries/scientific-realism/.

Chambers, R. G. 1960. “Shift of an Electron Interference Pattern by Enclosed Magnetic Flux.” *Physical Review Letters*. 5: 3-5.

Chen, Y. R., Wilczek, F., Witten, E., Halperin, B. I., 1989. “On Anyon Superconductivity.” *International Journal of Modern Physics B.* 3(7): 1001-1067.

Chomaz, P., Gulminelli, F., and Duflot, V. 2001. “Topology of Event Distributions as a Generalized Definition of Phase Transitions in Finite Systems.” *Physical Review E*, *64*(4), 046114.

Christensen, D. (1999). “Measuring Confirmation.” *Journal of Philosophy* 96: 437-461.

Clavelin, M. 1974. *The Natural Philosophy of Galileo*, translated by A. J. Pomeran. Cambridge: MIT Press.

Colyvan, M. 2001. *The Indispensability of Mathematics.* New York: Oxford University Press.

Colyvan, M, 2007. “Mathematical Recreation Versus Mathematical Knowledge.” In Leng, Paseau, and Potter (Eds.) *Mathematical Knowledge*. Oxford: Oxford University Press. 109–122.

Colyvan, M. 2010. “There is No Easy Road to Nominalism.” *Mind*. 119: 85–306.

Colyvan, M. 2012. “Road Work Ahead: Heavy Machinery on the Easy Road.” *Mind*. 121: 1031-1046.

Colyvan, M. 2015. “Indispensability Arguments in Philosophy of Mathematics.” In *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta (ed.) URL = http://plato.stanford.edu/entries/mathphil-indis/#2

Compagner, A. 1989. “Thermodynamics as the Continuum Limit of Statistical Mechanics.” *American Journal of Physics*, *57*(2), 106-117.

Contessa, G. 2007. “Scientific Representation, Interpretation and Surrogative Reasoning.” *Philosophy of Science*, 74, 48-68.

Contessa, G. 2011. “Scientific Models and Representation.” S. French and J. Saatsi (eds.) *The Continuum Companion To The Philosophy of Science*, 120.

da Costa, N. C. A. and French, S. 1990. “The Model-Theoretic Approach in the Philosophy of Science.” *Philosophy of Science* 57: 248-265.

da Costa, N. C. A. and French, S. 2003. *Science and Partial Truth*. Oxford: Oxford University Press.

Davey, K. 2011. "Idealizations and Cotenxtualism in Physics." *Philosophy of Science*. 78(1): 16-38.

Dizadji-Bahmani, F., Frigg, R., and Hartmann, S. 2010. "Who's Afraid of Nagelian Reduction?" *Erkenntnis*. 73 (3): 393-412.

Douçot, B., Pasquier, V., Duplantier, B., Rivasseau, V. (eds.) 2004. *The Quantum Hall Effect Poincaré Seminar*. Basel Boston Berlin: Birkhäuser Verlag.

Drake, S. (Translator) 1967. Galileo Galilei's *Dialogue Concerning the Two Chief World Systems: Ptolemaic and Copernican*. Berkeley: University of California Press.

Dresden, M. 1964. "The Existence and Signigicance of Parastatistics." In: Hayakawa, H. (ed.) *Lectures on Astrophysics and Weik Interactions*. Waltham, MA: Brandeis University. 377-469.

Duck, I., Sudarshan, E. C. G. (eds.) 1997. *Pauli and the Spin-statistics Theorem*. Singapore: World Scientific.

Earman, J. 1992. *Bayes or Bust?* Cambridge/MA: The MIT Press.

Earman, J. 2002. "Laws, Symmetry, and Symmetry Breaking; Invariance, Conservation Principles, and Objectivity." Manicrust. Available via http://philsciarchive.pitt.edu/archive/00000878/

Earman, J. 2003a. "The cosmological constant, the fate of the universe, unimodular gravity and all that." *Studies in History and Philosophy of Modern Physics*. 34(4): 559–577.

Earman, J. 2003b. "Rough Guide to Spontaneous Symmetry Breaking", in K. Brading and E. Castellani (eds.) *Symmetries in Physics: Philosophical Reflections*. Cambridge: Cambridge University Press, 334-345.

Earman, J. 2004. "Curie's Principle and Spontaneous Symmetry Breaking." *International Studies in the Philosophy of Science* 18: 173-198.

Earman, J. 2010. "Understanding permutation invariance in quantum mechanics." Unpublished manuscript.

Earman, J., Glymour, C., Mitchell, S. (eds.) (2002) *Ceteris Paribus Laws*. Netherlands: Kluwer Academic Publisher.

Ehrenberg, W., Siday, R. W. 1949. "The refractive index in electron optics and the principles of dynamics." *Proceedings of the Physical Society*. B62(1): 8-21.

Elgin, C. Z. 2004. "True enough." *Philosophical Issues.* (14): 113–131.

Ellis, B. 1992. "Idealizations in Science." In Dilworth C. (ed.) *Idealization IV: Intelligibility in Science*. Amsterdam, Netherlands: Rodopi.

Emch, G., Liu, C. 2002. *The Logic of Thermo-statistical Physics*. Berlin: Springer.

Ezawa, Z. F. 2013. *Quantum Hall Effects: Recent Theoretical and Experimental Developments*. River Edge, NJ: World Scientific.

Fadell, E., Neuwirth, L. 1962. "Configuration Spaces." *Mathematica Scandinavica*. 10: 111-118.

Fine, A. 1996. *The Shaky Game: Einstein, Realism and the Quantum Theory*. Second Edition. Chicago: The University of Chicago Press.

Fisher, R. A. 1930. *The Genetical Theory of Natural Selection*. Oxford: Clarendon.

Fox, R., Neuwirth, L. 1962. "The Braid Groups." *Mathematica Scandinavica*. 10: 119-126.

Frege, G. 1892. "On sense and reference." In G. Frege, P. Geach, & M. Black (Eds.), *Translations from the philosophical writings.* (pp. 56–78). Oxford: Basil Blackwell; 1960.

French, S. 2003. "A Model-Theoretic Account of Representation (Or, I Don't Know Much about Art … but I Know It Involves Isomorphism)" *Philosophy of Science*, *70*(5), 1472-1483.

French, S. 2014. *The Structure of the World: Metaphysics and Representation*. Oxford: Oxford University Press.

French, S. and Ladyman, J. 1999. "Reinflating the Semantic Approach." *International Studies in the Philosophy of Science*. 23:103-121.

van Fraassen, B. 1980. *The Scientific Image*. Oxford: Oxford University Press.

van Fraassen, B. 1989. *Laws and Symmetry*. Oxford: Oxford University Press.

Field, H. 1980. *Science Without Numbers: A Defense of Nominalism*. Oxford: Basil Blackwell.

Frigg, R. 2002. "Models and Representation: Why Structures Are Not Enough." *Measurement in Physics and Economics Project Discussion Paper Series*, DP MEAS 25/02. London School of Economics.

Frigg, R. 2006. "Scientific Representation and the Semantic View of Theories" *Theoria.* 50: 49-65.

Frigg, R., Hartmann, S. 2012. "Models in Science." In *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta (ed.) URL= http://plato.stanford.edu/entries/models-science/

Gale 1998. "Idealization in Cosmology: A Case Study" *Idealizations IX: Idealizations in Contemporary Physics*. Ed. N. Shanks. Poznan Studies in the Philosophy of the Sciences and the Humanities. Amsterdam: Rodopi, (1998): 165–182.

Gasiorowicz, S. 2003. *Quantum Physics*. John Wiley & Sons, Inc.

Gelfert, A. 2005. "Mathematical Rigor in Physics: Putting Exact Results in Their Place." Manucript. URL <http://philsci-archive.pitt.edu/2160/1/GELFERT_PSA2004_MathematicalRigor.pdf>

Giere, R. 1985. "Constructive Realism." in P.M. Churchland and C. Hooker (eds.), *Images of Science: Essays on Realism and Empiricism with a Reply from Bas C. can Fraassen*. Chicago: University of Chicago Press.

Giere, R. 1988. *Explaining Science: A Cognitive Approach*. Chicago: University of Chicago Press.

Giere, R. 1999a. *Science Without Laws*. Chicago: University of Chicago Press.

Giere, R. 1999b. "Using Models to Represent Reality." In L. Magnami, N. Nersessian, and P. Thagard (eds.) *Model-Based Reasoning and Scientific Discovery* pp. 41-57. New York: Kluwer/Plenum.

Giere, R. 2004. "How Models Are Used to Represent Reality." *Philosophy of Science* 71, Supplement, S742-S752.

Godfrey-Smith, P. 2006. "The Strategy of Model-Based Science." *Biology and Philosophy* 21: 725-740.

Goldstein, H., Poole, C. P. Jr., Safko, J. L. 2000. *Classical Mechanics*. San Francisco, USA: Adison Wesley.

Goodman, N. 1972. "Seven Strictures on Similarity" In *Problems and Projects*. Indianapolis: The Bobbs-Merrill Company.

Goodman, N. 1976. *Languages of Art: An Approach to a Theory of Symbols*, 2nd Edition. Indianapolis: Hackett Publishing Company.

Griffiths, D. J. 2005. *Introduction to Quantum Mechanics*. New Jersey, USA: Pearson Prentice Hall.

Gross, D. H. and Votyakov, E. V. 2000. "Phase Transitions in 'Small' Systems" *The European Physical Journal B-Condensed Matter and Complex Systems*, *15*(1), 115-126.

Grünbaum, A. 1954. “Science and Ideology.” *The Scientific Monthly* (July 1954).

Haldane, F. D. M. 1983. “Fractional Quantization of the Hall Effect: A Hierarchy of Incompressible Quantum Fluid States.” *Physics Review Letters*. 51: 605.

Hall, E. H. 1879. “On a New Action of the Magnet on Electric Currents.” *American Journal of Mathematics*. Vol. 2(3): 287–92.

Halperin, B. I. 1984. “Statistics of Quasiparticles and the Hierarchy of Fractional Quantized Hall States.” *Physical Review Letters*. 53: 1583-1586.

Hartmann, S. 1998. “Idealization in Quantum Field Theory.” In *Idealizations IX: Idealizations in Contemporary Physics*. Ed. N. Shanks. Poznan Studies in the Philosophy of the Sciences and the Humanities. Amsterdam: Rodopi, (1998): 99–122.

Hartmann, S. 1999. “Models and Stories in Hadron Physics.” In *Models as Mediators: Perspectives on Natural and Social Science*. Eds. M. S. Morgan and M. Morrison. Cambridge University Press, New York. (1999): 326–346.

Hartmann, S. and Sprenger, J. 2010. “Bayesian Epistemology.” In *Routledge Companion to Epistemology*. Eds. S. Bernecker and D. Pritchard. London: Routledge (2010): 609-620.

Hatcher, A. 2002. *Algebraic Topology*. Cambridge University Press.

Healey, R. A. 2007. *Gauging What’s Real: The Conceptual Foundations of Contemporary Gauge Theories.* New York: Oxford University Press.

Hempel, C. G and Oppenheim, P. (1948) “Studies in the Logic of Explanation.” *Philosophy of Science* 15: 135-75. Reprinted in *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, C. Hempel, New York: Free Press, (1965): 245–290.

Hempel, C. G. 1965. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press.

Hill, T. R. 1963. *Thermodynamics of Small Systems: Part I*. New York: W. A. Benjamin, INC. Publishers.

Hohensee, M. A., Estey, B., Hamilton, P., Zeilinger, A., and Müller, H. 2012. “Force-Free Gravitational Redshift: Proposed Gravitational Aharonov-Bohm Experiment.” *Physical Review Letters*. 108: 230404.

Hooker, C. A. 1994. “Idealization, Naturalism, and Rationality: Some Lessons from Minimal Rationality” *Synthese* 99:181-231.

Hopkins, R. 2005. “The speaking image: Visual communication and the nature of depiction.” In M. Kieran (Ed.), *Contemporary debates in aesthetics and the philosophy of art*. New York: Blackwell.

Howson, C. and Urbach, P. 1993. *Scientic Reasoning: The Bayesian Approach*. La Salle: Open Court.

Huang, W.-H. 2003. “Boson-Fermion Transmutation and the Statistics of Anyons.” Available via arXiv:hep-th/0308095

Hughes, R .I.G. 1990. “The Bohr Atom, Models and Realism.” *Philosophical Topics* 18:71-84.

Hughes, R.I.G. 1993. “Theoretical Explanation” in French, P. (eds.) (1993) *Midwest Studies in Philosophy, Volume XVIII, Philosophy of Science.*

Hughes, R.I.G. 1997. “Models and Representations.” *Philosophy of Science* 64: S325-S336.

Jackiw, R. 1990. “Dynamical symmetry of the magnetic vortex.” *Annals Of Physics*. 201: 83-116.

Jackiw, R., Tempelton, S. 1981. “How Super-renormalizable Interactions Cure their Infrared Divergences.” *Physical Review D*. **23**: 2291.

Jacob, P. 2010. “Intentionality.” *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2012/entries/davidson/>.

Jones, N. J. 2006. *Ineliminable Idealizations, Phase Transitions and Irreversibility*. Dissertation, Ohio State University.

Jones, M. R. 2005. “Idealization and Abstraction: A Framework.” In Martin R. Jones and Nancy Cartwright. (eds) *Idealization XII: Correcting the Model.* (2005) New York: Rodopi.

Joyce, J. 1999. *The Foundations of Causal Decision Theory*. Cambridge: Cambridge University Press.

Kadanoff, L. P. 2000. *Statistical Physics: Statics, Dynamics and Renormalization*. Singapore: World Scientific.

Kaplan, I. G. 1994. “Symmetry of Many-Particle Systems.” In: Bonchev, D., Rouvray, D. H. (eds.) Chemical Group Theory: Introduction and Fundamentals. Amsterdam: Gordon and Breach Science Publishers. 209-254.

Katanaev, M. O. 2011. “On Geometric Interpretation of the Aharonov-Bohm Effect.” *Russian Physics Journal*. 54(5): 507-514.

Katanaev, M. O. 2012. “On Geometric Interpretation of the Berry Phase.” *Russian Physics Journal*. 54(10): 1082-1092.

Khare, A. 2005. *Fractional Statistics and Quantum Theory*. NJ, USA: World Scientific.

Khlentzos, D. 2011 “Challenges to Metaphysical Realism.” In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. URL <http://plato.stanford.edu/entries/realism-sem-challenge/.Kim, J. 1999. “Making Sense of Emergence.” *Philosophical Studies.* 95 (1-2):3-36.

Kitcher, P. 1981. “Explanatory Unification.” *Philosophy of Science*, 48: 507-531.

Kitcher, P. 1989. “Explanatory Unification and the Causal Structure of the World.” In Kitcher, P. and Salmon, W. (eds.) *Minnesota Studies in the Philosophy of Science, Vol. XIII: Scientific Explanation*. Minneapolis: University of Minnesota Press, pp. 410-505.

Kittel, C. 2005. *Introduction to Solid State Physics*. John Wiley & Sons, Inc.

von Klitzing, K. 2004. “25 Years of Quantum Hall Effect (QHE): A Personal View on the Discovery. Physics and Application of this Quantum Effect.” In: Douçot, B., Pasquier, V., Duplantier, B., Rivasseau, V. (eds.) *The Quantum Hall Effect Poincaré Semina.* Vol. 45: 1-23. Basel Boston Berlin: Birkhäuser Verlag.

von Klitzing, K., Dorda, G., and Pepper, M. 1980. "New Method for High-Accuracy Determination of the Fine-Structure Constant Based on Quantized Hall Resistance". *Physical Review Letters*. 45:494-497.

Koertge, N. 1977. “Galileo and the Problem of Accidents” *Jounal of the History of Ideas.* 389-408.

Koperski, J. 2006. “Models.” *The Internet Encyclopedia of Philosophy*, ISSN 2161-0002, http://www.iep.utm.edu/models/

Kuhn, T. 1970. *The Structure of Scientific Revolutions*. Second Edition. Chicago: University of Chicago Press.

Kuhn, T. 1983. “Commensurability, Comparability, Communicability.” P.D. Asquith & T. Nickles (eds.), *Proceedings of the Philosophy of Science Association, 1982*: 669–688.

Kuper, C. G. 1980. “Electromagnetic Potentials in Quantum Mechanics: A Proposed Text of the Aharonov-Bohm Effect.” *Physics Letters* 79A: 413-416.

Krajewski, W. 1997. *Correspondence Principle and the Growth of Science*. Dordrecht: Reidel.

Krauss, L. M. 1994. *Fear of Physics: A Guide For The Perplexed*. Jonathan Cape.

Kretzschmar, M. 1965. “Aharonov-Bohm Scattering of a Wave Packet of Finite Extension.” *Zeitschrift für Physik.* 185: 84–96.

Ladyman, J. 2008. “Idealization” *The Routledge Companion to Philosophy of Science*. Eds. Psillos, S. and Curd, M. London and New York: Routledge Taylor & Francis Group. (2008): 358-366.

Laidlaw, M. G., DeWitt, C. M. 1971. "Feyman Functional Integrals for System of Indistinguishable Particles." *Physical Review D*. **3:** 1375-1378.

Landsman, N. P. 2013. "Quantization and superselection sectors III: Multiply connected spaces and indistinguishable particles." Available via arXiv:1302.3637

Laughlin, R. B. 1983a. "Quantized Motion of Three Two-dimensional Electrons in a Strong Magnetic Field." *Physical Review B*. 27: 3383-3389.

Laughlin, R. B. 1983b. "Anomalous Quantum Hall Effect: An Incompressible Quantum Fluid with Fractionally Charged Excitations." *Physical Review Letters*. 50: 1395-1398.

Laughlin, R. B. 1990. "Fractional Statistics in the Quantum Hall Effect." In: Wilczek, F. (ed.): *Fractional Statistics and Anyon Superconductivity*. Singapore: World Scientific. 262-306.

Laymon, R. 1980. "Idealization, Explanation, and Confirmation." in *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association, Vol. I* pp. 336-350.

Laymon, R. 1984. "The Path from Data to Theory." In *Scientific Realism* J. Leplin (ed.) Berkeley: University of California Press, 108-123.

Laymon, R. 1985. "Idealizations and the Testing of Theories by Experimentation." In Peter Achinstein and Owen Hannaway (eds.), *Observation Experiment and Hypothesis in Modern Physical Science*. Cambridge, Mass.: M.I.T. Press, 147-173.

Laymon, R. 1989a. "Applying Idealized Scientific Theories to Engineering." *Synthese*, 81:353-371.

Laymon, R. 1989b. "Cartwright and the Lying Laws of Physics." *The Journal of Philosophy* 86(7): 353-372.

Laymon, R. 1995. "Experiment and the Legitimacy of Idealization." *Philosophical Studies*, 77: 353-375.

Lebowitz, J. L. 1999. "Statistical Mechanics: A Selective Review of Two Central Issues." *Reviews of Modern Physics* 71(2):S346-S357.

Leeds, S. 1999. "Gauges: Aharonov, Bohm, Yang, Healey." *Philosophy of Science.* 66: 606–627.

Leinaas, J. M., Myrheim, J. 1977. „On the Theory of Identical Particles." *Nuovo Cimento B*. **37:** 1-23.

Leng, M. 2010. *Mathematics and Reality*. Oxford: Oxford University Press.

Leng, M. 2012. "Taking it Easy: A Response to Colyvan." *Mind.* 121: 983–96.

Lerda, A. 1992. *Anyons: Quantum Mechanics of Particles with Fractional Statistics*. Berlin Heidelberg: Springer-Verlag.

Levins, R. 1966. "The Strategy of Model Building in Population Biology." In E. Sober (Ed.) *Conceptual Issues in Evolutionary Biology*. Cambridge, MA: MIT Press.

Lewis, D. 1970. "General Semantics," *Synthese* 22: 18–67.

Lewis, D. 1986. "Causal Explanation." In *Philosophical Papers, Vol. II*. New York: Oxford University Press, 214-240.

Liggins, D. 2012. "Weaseling and the Content of Science." *Mind*. 121: 997–1006.

Lind, H. 1993. "A Note on Fundamental Theory and Idealizations in Economics and Physics." *British Journal for Philosophy of Science* 44: 493-503.

Liu, C. 1999a. "Approximation, Idealization, and the Laws of Nature." *Synthese* 118: 229-256.

Liu, C. 1999b. "Explaining the Emergence of Cooperative Phenomena." *Philosophy of Science* 66 (Proceedings): S92–S106.

Liu, C. 2001. "Infinite Systems in SM Explanations: Thermodynamic Limit, Renormalization (Semi-)Groups, and Irreversibility" *Philosophy of Science Proceedings*, 68:S325-S344.

Liu, C. 2004a. "Approximations, Idealizations and Models in Statistical Mechanics" *Erkenntnis*, *60*(2), 235-263.

Liu, C. 2004b. "Laws and Models in a Theory of Idealization," *Synthese* 138: 363-385.

Liu, C. 2007. "Confirmation Idealized Theories and Scientific Realism." in http://philsci-archive.pitt.edu/3244/

Liu, C., Emch, G. 2005. "Explaining quantum spontaneous symmetry breaking." *Studies in History and Philosophy of Modern Physics*. 36: 137–164.

Lloyd, E. A. 1994. *The Structure and Confirmation of Evolutionary Theory*. (2nd ed.) Princeton: Princeton University Press.

Lyre, H. 2001. "The Principles of Gauging." *Philosophy of Science* 68(3):S371–S381.

Lyre, H. 2004. "Holism and Structuralism in U(1) Gauge Theory." *Studies in History and Philosophy of Modern Physics*. 35: 643-670.

Magni, C. and Valz-Gris, F. 1995. "Can Elementary Quantum Mechanics Explain the Aharonov-Bohm effect?" J*ournal of Mathematical Physics*. 36(1): 177-186.

Mäki, U. 1994. "Isolation, Idealization and Truth in Economics" in Hamminga, B. and De Marchi, N. B. (eds.), *Idealization VI: Idealization in Economics*. Poznan Studies in the Philosophy of the Sciences and the Humanities, Vol. 38: 147–168. Amsterdam: Rodopi.

Mainwood, P. R. 2006. *Is More Different? Emergent Properties in Physics*. PhD Dissertation, Oxford University.

Mathewson, J. and Weisberg, M. 2009. "The Structure of Tradeoffs in Model Building." Synthese. 170(1): 169-190.

Mattingly, J. 2006. "Which Gauge Matters." *Studies in History and Philosophy of Modern Physics*. 37: 243-262.

Maudlin, T. 1998. "Discussion: Healey on the Aharonov–Bohm Effect." *Philosophy of Science*. 65: 361–368.

Maxwell, J. C. 1892. *A Treatise on Electricity and Magnetism.* Vol. II (Third Edition) Oxford at The Clarendon Press.

McMullin, E. 1985. "Galilean Idealization." *Studies In History And Philosophy of Science*, 16 (3):247–273.

Melia, J. 2000. "Weaseling Away the Indispensability Argument." *Mind*. 109: 455–7.

Mendell, H. 2004. "Aristotle and MAthematics." In *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta (ed.) URL= http://plato.stanford.edu/entries/aristotle-mathematics/

Menon, T., and Callender, C. 2013. "Turn and Face the Strange… Ch-ch-changes: Philosophical Questions Raised by Phase Transitions." In *The Oxford Handbook of Philosophy of Physics*. R. Batterman (ed) Oxford: Oxford University Press. (2013): 189-223.

Messiah, A. M., Greenberg, O. W. 1964. "Symmetrization Postulate and its Experimental Foundation." *Physical Review B*. 136: 248-267.

Miller, A. 2010. "Realism". In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. URL=http://plato.stanford.edu/entries/realism/.

Miller, J. H. and Page, S. E. 2007. *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*. Princeton: Princeton University Press.

Milne, P. 1996. "log[p(h/eb)/p(h/b)] is the one true measure of confirmation." *Philosophy of Science* 63: 21-26.

Minkowski, H. 1908. "Space and time." In H. A. Lorentz, A. Einstein, H. Minkowski, & H. Weyl (Eds.), *Principle of relativity*. (pp. 75–91). New York: Dover; 1952.

Morandi, G. 1992. *The Role of Topology in Classical and Quantum Mechanics*. Berlin: Springer-Verlag.

Morandi, G., Menossie, E. 1984.: "Path-integrals in Multiply Connected Spaces and the Aharonov-Bohm Effect." *European Journal of Physics* 5:49-58.

Morgan, M. and M. Morrison. 1999. *Models as Mediators. Perspectives on Natural and Social Science*. Cambridge: Cambridge University Press.

Morrison, M. 1999. "Models as Autonomous Agents." in Morgan and Morrison (eds), 38-65.

Morrison, M. 2000. *Unifying Scientific Theories*. Cambridge: Cambridge University Press.

Morrison, M. 2006. "Emergence, Reduction and Theoretical Principles: Rethinking Fundamentalism." *Philosophy of Science*. 73: 876–887.

Morrison, M. 2012. "Emergent Physics and Micro-Ontology" *Philosophy of Science* 79:141-166.

Mourik, V., Zuo, K., Frolov, S. M., Plissard, S. R., Bakkers, E. P. A. M., Kouwenhoven, L. P. 2012. "Signatures of Majorana Fermions in Hybrid Superconductor-Semiconductor Nanowire Devices." *Science*. 336: 1003-1007.

Mundy, B. 1986. "On the General Theory of Meaningful Representation." *Synthese* 67:391-437.

Munkres, J. R. 2000. *Topology*. NJ, USA: Prentice Hall.

Musgrave, A. 1981. "'Unreal Assumptions' in Economic Theory: The F-Twist Untwisted", *Kyklos* 34: 377-387.

Nagel, E. 1961 *The Structure of Science*. Routledge and Kegan Paul, London.

Nakhara, M. 1990. *Geometry, Topology and Physics*. Bristol: Institute of Physics Publishing.

Nash, C., Sen, S. 1983. *Topology and Geometry for Physicists*. New York: Academic.

Nayak, C., Simon, S., Stern, A., Freedman, M., Das Sarma, S. 2008. "Non-Abelian anyons and topological quantum computation." *Reviews of Modern Physics*. 80: 1083.

Niiniluoto, I. 1990. "Theories, Approximations and Idealizations. In J. Brezeziński, F. Coniglione, et al. (eds) *Idealization I: General Problems*. Amsterdam: Rodopi. 9-57.

Niiniluoto, I. 1999. *Critical Scientific Realism*. Oxford University Press.

Norton, J. D. 2011. "Analogy." Manuscript. Available via http://www.pitt.edu/~jdnorton/papers/Analogy.pdf

Norton, J. D. 2012. “Approximations and Idealizations: Why the Difference Matters.” *Philosophy of Science* 79, 207–32.

Nounou, A. M. 2003. “A Fourth Way to the Aharonov-Bohm Effect.” In: Bradind, K., Castellani, E. (eds.) *Symmetries in Physics: Philosophical Reflections.* Cambridge University Press.

Nowak, L. 1980. *The Structure of Idealization*. Dordrecht: Reidel.

Nowak, L. 1989. “On the (Idealizational) Structure of Economic Theories.” *Erkenntnis*, 30, 225-246.

Nowak, L. 1991. “Thoughts are Facts of Possible Worlds. Truths are Facts of a Given World.” *Dialectica* 45: 273-287.

Nowak, L. 1992. “The Idealizational Approach to Science: A Survey” in Brzezinski, J. and Nowak, L. (eds.) 1992. *Idealization III: Approximation and Truth*. Amsterdam: Rodopi.

Nowakowa, I. and Nowak, L. (eds.). 2000. *Idealization X: The Richness of Idealizations*. Amsterdam: Rodopi.

Oddie, G. 2007. “Truthlikeness.” *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), URL = http://plato.stanford.edu/entries/truthlikeness/

de Oliveira, C. R., M. Pereira. 2008. “Mathematical Justification of the Aharonov-Bohm Hamiltonian.” *Journal of Statistical Physics*. 133:1175-1184.

de Oliveira, C. R., M. Pereira. 2010. “Scattering and Self-adjoint extensions of the Aharonov-Bohm Hamiltonian.” *Journal of Physics A: Mathematical and Theoretical*. 43:1-29.

de Oliveira, C. R., M. Pereira. 2011. “Impenetrability of Aharonov-Bohm Solenoids: Proof of Norm Resolvent Convergence.” *Letters in Mathematical Physic.s* 95: 41-51.

O'Neill, O. 2000. *Bounds of Justice*. New York: Cambridge University Press.

Pachos, J. K. 2012. *Introduction to Topological Quantum Computation*. NY, USA: Cambridge University Press.

Pashby, T. (2012). “Dirac’s prediction of the positron: A case study for the current realism debate.” *Perspectives on Science,* 20(4): 440–475.

Pauli, W. 1940. “The connection between spin and statistics.” *Physical Review*. 58: 716–722.

Peebles, P. J. E. 1992. *Quantum Mechanics*. NJ, USA: Princeton University Press.

Peirce, C. S. 1998. *The Essential Peirce: Selected Philosophical Writings.* Second Edition. N. Houser, C. Kloesel, & the Peirce Edition Project (eds.). Bloomington: Indiana University Press.

Peshkin, M., Tonomura, A. 1989. *The Aharonov-Bohm Effect*. LNP, vol. 340. Berlin: Springer.

Pincock, C. 2012. *Mathematics and Scientific Representation*. Oxford University Press.

Prange, R.E., Girvin, S. (eds.) 1987. *The Quantum Hall Effect*. New York: Springer.

Prigogine, I. 1997. *The End of Certainty*. New York: The Free Press.

Psillos, S. 1999. *Scientific Realism: How Science Tracks Truth*. Routledge.

Putnam, H. 1971. *Philosophy of Logic*. New York: Harper.

Putnam, H. 1975. *Mathematics, Matter and Method*. Cambridge: Cambridge University Press.

Quine, W. V. O. 1960. *Word and Object*. Cambridge, MA: M.I.T. Press.

Quine, W. V. O. 1981. *Theories and Things*. Cambridge, MA: Harvard University Press.

Rao, S. 2001. "An Anyon Primer." Manuscript. Available via arXiv:hep-th/9209066

Reed, M., Simon, B. 1980. *Methods of Modern Mathematical Physics I: Functional Analysis*. San Diego, CA: Academic Press, Inc.

Reimer, M. 2009. "Reference." In *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta (ed.) URL= http://plato.stanford.edu/entries/reference/

Rosen, G. 2001. "Abstract Objects." *Stanford Encyclopedia of Philosophy*. Edward N Zalta (ed.) URL= http://plato.stanford.edu/entries/abstract-objects/

Rosenkrantz, R. D. 1994. "Bayesian Confirmation: Paradise Regained." *British Journal for the Philosophy of Science* 45: 467-476.

Roy, S.M. 1980. "Condition for Nonexistence of Aharonov-Bohm Effect." *Physical Review Letters*. 44: 111–114.

Rueger, A. and Sharp, D. W. (1998) "Idealization and Stability: A Perspective from Nonlinear Dynamics." In Shanks, N. (ed.) *Idealization IX: Idealization in Contemporary Physics*. Amsterdam-Atlanta, GA: Rodopi. pp. 201-216.

Ruetsche, L. 2011. *Interpreting Quantum Theories: The Art of the Possible*. Oxford: Oxford University Press.

Ryder, L. H. 1996. *Quantum Field Theory*. Cambridge : Cambridge University Press.

Salmon, W. C., 1998. *Causality and Explanation*. Oxford University Press: New York.

Schulman, L. 1971. “A Path Integral for Spin.” *Physical Review* 176:1558-1569.

Sellars, W. 1963. *Science, Perception and Reality.* New York: The Humanities Press.

Sewell, G. L. 1986. *Quantum Theory of Collective Phenomena*. Oxford University Press.

Sewell, G. L. 2002. *Quantum Mechanics and its Emergent Macrophysics*. Princeton University Press.

Shaffer, M. J. 2001. “Bayesian Confirmation of Theories that Incorporate Idealizations.” *Philosophy of Science* 66: 36-52.

Shaffer, M. J. 2012. *Counterfactuals and Scientific Realism*. UK: Palgrave Macmillan.

Shapere, A., Wilczek, F. (eds.) 1989. *Geometric Phases in Physics*. Singapore: World Scientific.

Shrivastava, K. N. 2005. *Quantum Hall Effect: Expressions*. New York: Nova Science Publishers.

Sider, T. 2011. Writing the Book of the World. Oxford: Oxford University Press.

Siewert, C. 2006. “Consciousness and Intentionality.” In *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta (ed.) URL= http://plato.stanford.edu/entries/consciousness-intentionality/

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

Sklar, L. M. 2000. *Theory and Truth: Philosophical Critique within Foundational Science*. Oxford: Oxford University Press.

Sklar, L. M. 2003. “Dappled Theories in a Uniform World.” *Philosophy of Science* 70:424-441.

Speaks, J. 2014. “Theories of Meaning.” In *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta (ed.) URL= http://plato.stanford.edu/entries/meaning/

Stanley, H. E. (1971) *Introduction to Phase Transitions and Critical Phenomena*. Oxford University Press.

Stern, A. 2008. “Anyons and the Quantum Hall effect—A Pedagogical Review.” Annals Of Physics. 323(1): 204-249.

Stich, S., & Warfield, T. A. 1994. *Mental representation: A reader*. Oxford: Blackwell.

Strevens, M. 2004 “The Causal and Unification Accounts of Explanation Unified—Causally.” *Nous*, XXXVIII, pp. 154-176.

Strevens, M. 2009. *Depth: An Account of Scientific Explanation*. Harvard University Press.

Styer, D. F. 2004. “What Good is the Thermodynamic Limit?” *American Journal of Physics* 72(1):25-29.

Suárez, M. 2003. “Scientific Representation: Against Similarity and Isomorphism.” *International Studies in the Philosophy of Science* 17(3).

Suárez, M. 2004. “An Inferential Conception of Scientific Representation” *Proceedings of the Philosophy of Science Association* 71(5):767-779.

Suárez, M. 2010. “Scientific Representation.” *Philosophy Compass*. *5*(1): 91–101.

Suárez, M. 1999. “Theories, Models and Representations” *Model-Based Reasoning in Scientific Discovery*, 75-83. Springer US.

Suppes, P. 1967. “What is a Scientific Theory?” in S. Morgenbesser (ed.) *Philosophy of Science Today*, New York: Basik Books, pp. 55-67.

Swoyer, C. 1991. “Structural Representation and Surrogative Reasoning.” *Synthese* 87: 449–508.

Teller, P. 1979. “Quantum Mechanics and the Nature of Continuous Physical Quantities.” *Journal of Philosophy*. 345–361.

Teller, P. 1995. *An Interpretive Introduction to Quantum Field Theory*. Princeton, New Jersey: Princeton University Press.

Teller, P. 2001. “Twilight of the Perfect Model” *Erkenntnis*. 55, 393–415.

Teller, P. 2004. “How We Dapple the World” *Philosophy of Science*. 71: 425–447.

Teller, P. 2011. “Two Models of Truth.” *Analysis*. 71: 465-472.

Tonomura, A. 1999. *Electron Holography*. Berlin Heidelberg New York: Springer-Verlag.

Tonomura, A. 2010. “The AB Effect and its Expanding Applications.” Journal of Physics A: Mathematical and Theoretical 43(35): 1-13.

Tonomura, A. Matsuda, T., Suzuki, R., Fukuhara, A., Osakabe, N., Umezaki, H., Endo, J., Shinagawa, K., Sugita, Y., Fujiwara., H. 1982. “Observation of Aharonov-Bohm effect by electron holography.” *Physical ReviewLetters*. 46: 1443.

Tonomura, A., Osakabe, N. , Matsuda T., Kawasaki T., Endo J.,Yano S., Yamada., H. 1986. "Evidence for Aharonov-Bohm effect with magnetic field completely shielded from electron wave." *Physical ReviewLetters*. 56: 792-795.

Tsui, D.C., Stormer, H.L., Gossard, A.C. 1982. "Two-Dimensional Magnetotransport in the Extreme Quantum Limit." *Physical Review Letters*. 48 (22): 1559.

Warfield, T. A., & Stich, S. 1994. "Introduction." In Stich, S., & Warfield, T. A. 1994. *Mental representation: A reader*. Oxford: Blackwell. 1-7.

Wayne, A. 2009. "Emergence and Singular Limits." *Synthese* 184(3):341-356.

Weisberg, M. 2006. "Forty Years of 'The Strategy': Levins on Model Building and Idealization." *Biology and Philosophy*. 21(5): 623-645.

Weisberg, M. 2007. "Three Kinds of Idealizations." *The Journal of Philosophy* 104 (12) 639-59.

Weisberg, M. 2013. *Simulation and Similarity: Using Models to Understand the World*. New York: Oxford University Press.

Weisskopf, V.F. 1961. "Selected Topics in Theoretical Physics." In Brittin, W.E. (ed.) *Lectures in Theoretical Physics. Vol. III*: New York: Interscience. 67–70.

Wilczek, F. 1982a. "Magnetic Flux, Angular Momentum and Statistics." *Physical Review Letters*. 48: 1144-1146.

Wilczek, F. 1982b. "Quantum Mechanics of Fractional-Spin Particles." *Physical Review Letters*. 49: 957-959.

Wilczek, F. (ed.) 1990. *Fractional Statistics and Anyon Superconductivity*. Singapore: World Scientific.

Wimsatt, W. C. 2007. *Re-engineering Philosophy for Limited Beings*. Cambridge, MA: Harvard University Press.

Wollheim, R. 1987. *Painting as an art*. London and New York: Thames and Hudson.

Woodward, J. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

Wu, T. T., C. N. Yang . 1975. "Concept of Nonintegrable Phase Factors and Global Formulation of Gauge Fields." *Physical Review*. D12: 3845.

Wu, Y. S. 1984. "General Theory for Quantum Statistics in Two Dimensions." *Physical Review Letters*. 52: 2103-2106.

Yablo, S. 2012. “Explanation, Extrapolation, and Existence.” *Mind*. 121: 1007–30.

Yoshioka, D. 2002. *Quantum Hall Effect*. Berlin: Springer.

Zetelli, N. 2009. *Quantum Mechanics: Concepts and Applications*. John Wiley and Sons, Ltd. West Sussex, UK.