INTERVENTIONIST CAUSATION IN PHYSICAL SCIENCE

by

Karen R. Zwier

B.S., Computer Engineering, University of Illinois, 2006

B.A., Philosophy, University of Illinois, 2006

M.A., Philosophy, University of Pittsburgh, 2011

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

University of Pittsburgh

2014

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Karen R. Zwier

It was defended on

November 20, 2014

and approved by

Sandra Mitchell, University of Pittsburgh, HPS

John Norton, University of Pittsburgh, HPS

James Woodward, University of Pittsburgh, HPS

Robert Batterman, University of Pittsburgh, Philosophy

Mark Wilson, University of Pittsburgh, Philosophy

Dissertation Director: Sandra Mitchell, University of Pittsburgh, HPS

INTERVENTIONIST CAUSATION IN PHYSICAL SCIENCE

Karen R. Zwier, PhD

University of Pittsburgh, 2014

The current consensus view of causation in physics, as commonly held by scientists and philosophers, has several serious problems. It fails to provide an epistemology for the causal knowledge that it claims physics to possess; it is inapplicable in a prominent area of physics (classical thermodynamics); and it is difficult to reconcile with our everyday use of causal concepts and claims.

In this dissertation, I use historical examples and philosophical arguments to show that the interventionist account of causation constitutes a promising alternative for a "physically respectable" account of causation. The interventionist account explicates important parts of the experimental practice of physics and important aspects of the ways in which physical theory is used and applied. Moreover, the interventionist account succeeds where the consensus view of causation in physics fails.

I argue that the interventionist account provides an epistemology of causal knowledge in physics that is rooted in experiment. On the interventionist view, there is a close link between experiment and the testing of causal claims. I give several examples of experiments from the early history of thermodynamics that scientists used in interventionist-type arguments. I also argue that interventionist claims made in the context of a physical theory can be epistemically justified by reference to the experimental interventions and observations that serve as evidence for the theory.

I then show that the interventionist account of causation is well-suited to the patterns of reasoning that are intrinsic to thermodynamic theory. I argue that interventionist reasoning

constitutes the structural foundation of thermodynamic theory, and that thermodynamic theory can provide clear answers to meaningful questions about whether or not a certain variable is a cause of another in a given context.

Finally, I argue that the interventionist account offers the prospect of a unification of "physically respectable" causation and our everyday notion of causation. I conclude the dissertation by sketching an anti-foundationalist unification of causation, according to which causal reasoning occurs in the same manner in physics as it does in other branches of life and scientific research.

TABLE OF CONTENTS

PREFACE . . . ix
1.0 INTRODUCTION . . . 1
2.0 EXPERIMENT AS A TEST OF CAUSAL CLAIMS: A HISTORY . . . 10
2.1 Background: Experience as evidence . . . 12
2.2 Galileo Galilei . . . 14
2.2.1 Galileo's devised experiences to determine a cause . . . 17
2.2.2 Galileo's notion of cause . . . 24
2.3 Francis Bacon . . . 35
2.3.1 Bacon's method . . . 38
2.3.2 Baconian methodology: Innovations and limitations . . . 41
2.4 John Stuart Mill . . . 44
2.4.1 Mill's two distinct concepts of causation . . . 45
2.4.2 Sorting out the two concepts . . . 48
2.5 Conclusion . . . 50
3.0 CAUSAL INFERENCE FROM EXPERIMENT: AN EPISTEMOLOGY 51
3.1 The interventionist account of causation . . . 53
3.1.1 Defining cause in terms of hypothetical experiment . . . 56
3.1.2 Ideal intervention . . . 60
3.2 From hypothetical experiment to real experiment . . . 65
3.2.1 The "Fundamental Problem" of causal inference . . . 66
3.2.2 Non-ideal interventions . . . 70
3.3 From real experiments to causal claims . . . 72

3.3.1 Example: The Berti experiment . . . 73
3.3.2 When does an experiment afford causal inference? . . . 80
4.0 THE EARLY HISTORY OF THERMODYNAMICS: A CASE STUDY 82
4.1 Torricelli's experiments . . . 83
4.2 Roberval's experiments . . . 92
4.3 Pascal's experiments . . . 98
4.4 Boyle's experiments and theorizing . . . 104
4.4.1 Boyle's J-tube experiment . . . 105
4.4.2 First-order abstraction . . . 108
4.4.3 Second-order abstraction . . . 111
4.4.4 Boyle's raised tube experiment . . . 116
4.5 Physical theory: More than an equation . . . 119
4.6 Conclusion . . . 123
5.0 INTERVENTIONIST CAUSATION IN PHYSICAL THEORY . . . 124
5.1 Clausius' memoirs, 1850–1865 . . . 126
5.1.1 1850 paper . . . 126
5.1.2 1854 paper . . . 133
5.1.3 1865 paper . . . 139
5.2 Interventionist causal reasoning in thermodynamics . . . 142
5.2.1 The Clausius submanifold . . . 142
5.2.2 Thermodynamic potentials and driving forces . . . 150
5.3 Conclusion . . . 158
6.0 PROSPECTS FOR A UNIFICATION OF CAUSATION . . . 159
6.1 Prospects for interventionist causation in other subdomains of physics . . . 161
6.2 A unified concept of causation . . . 163
6.2.1 Anti-foundationalist unification . . . 166
6.2.2 Anti-reductionist unification . . . 168
6.3 Conclusions and future research . . . 171
BIBLIOGRAPHY . . . 173

LIST OF FIGURES

2.1 Galileo's illustration of a device for measuring the "force of the void" 23
2.2 Galileo's illustration of two cylindrical beams of differing dimensions 29
2.3 Galileo's illustration of why an ebony chip floats on the surface of water . . . 31
3.1 Illustration of multiple instances of an experiment 68
3.2 Diagram of Berti's experiment 75
3.3 Engraving of a more complex version of Berti's experiment 76
4.1 Diagram of Torricelli's experiment 86
4.2 Diagram of Torricelli's thought experiment about compressible wool 91
4.3 Diagram of Roberval's experiment with the carp bladder 95
4.4 Illustration of one type of vacuum-in-a-vacuum experiment 98
4.5 Boyle's data table for the J-tube experiment 107
4.6 Boyle's data table for the raised tube experiment 117
5.1 The 4-stage thermodynamic cycle in Clausius' 1850 paper 129
5.2 The 6-stage thermodynamic cycle in Clausius' 1854 paper. 135
5.3 Illustration of the Clausius submanifold 145
5.4 Two thermodynamic systems finding an equilibrium temperature 151
5.5 Illustration of a pressure-driven process. 155

PREFACE

This dissertation is a labor of several years and I have many people to acknowledge for their support.

First of all, I thank the members of my committee. I consider myself fortunate to have benefited from their knowledge, experience, and goodwill. Each of them has helped me to clarify and build the ideas and arguments contained here, and more holistically speaking, each of them has influenced me as a scholar. Bob Batterman managed to steer me away from a naïve reductionist argument that I was attempting to make at one point in the process. John Norton was always quick to respond and make time for conversation. Those conversations often provided the intellectual stimulus and motivation that I needed. Mark Wilson has shown me, with his unique style, that there are many different ways of doing valuable philosophical work. His ideas and his ways of expressing them are unlike those of any other philosopher I have encountered, and he has given me so much to ruminate on. I was lucky enough to have Jim Woodward come to the University of Pittsburgh half-way through my graduate career, after I had already come to admire his philosophical ideas. I thank him for his thorough feedback on my writing. My advisor, Sandy Mitchell, saw both the good ideas and the bad. She had a clear eye for what to discard (the pages of which are probably about twice as long as the finished version) and how to point me in fruitful directions. I thank her especially for her support of me at every stage. For a woman to give birth to three children while working toward her doctorate is neither a standard nor an efficient path through the endeavor. Yet Sandy was unceasingly supportive of my life choices and was a constant advocate for me professionally. She played the midwife to this dissertation, my fourth "baby".

I would additionally like to thank Jim Lennox. He has been a wonderful teacher to me

in historical and philosophical domains not covered in this dissertation. I would like to think that I get my integrated-HPS style from him. Thanks also to Paolo Palmieri and Peter Machamer for their comments on chapter 2.

It has been a great joy to be a part of the Department of History and Philosophy of Science at the University of Pittsburgh. The place is a professional utopia for historians and philosophers of science and I have been so enriched in my eight-and-a-half years as a student here. I thank the graduate student community in particular. It's hard to imagine how a group of people could be more friendly, fun, intelligent, motivated, and interestingly wacky. I am so very grateful to STARS: Julia Bursten, Bihui Li, Elizabeth O'Neill, Aleta Quinn, Catherine Stinson, and Katie Tabb. They have helped me build my confidence and repeatedly reminded me not to undersell myself. Many thanks to Meghan Dupree and Kathryn Lindeman for being my sisters in philosophy and in Christ. And finally, I feel the deepest gratitude toward Peter Distelzweig and Jonah Schupbach, for their friendship and fellowship and many edifying conversations. I have been blessed to have had them both in my cohort and to have walked the graduate school journey with them.

To all the people, known and unknown, who have created the many good writing places that I have utilized, I offer my thanks. Special thanks go to Suzan Erem and Paul Durrenberger for their wonderful hospitality at Draco Hill Writers' Retreat, where I managed to get over the final hump in writing this dissertation.

Thanks to Michelle Sherman and her family for not only caring for my children while I worked but enriching them as well.

To Mom, Dad, Craig, Carl, and Drew: thank you for all the love you gave me during my idyllic childhood. Thank you especially for encouraging and supporting me despite my exasperating habit of taking on overly ambitious projects (a PhD, for example).

To Claire Lucia, Joshua Cecil, and Benjamin Jude: I love you so very much. Thank you for the joy you bring to my days. Despite my efforts to prevent my work from disrupting your lives, you still have had to tolerate a lot of upheaval and a more-than-occasionally frazzled mother. I hope that where I have failed in my efforts to provide perfect stability I have succeeded in inspiring in you curiosity, engagement, and a sense of vocation. And Claire, if you do truly want to be "Dr. Z 3" someday, I believe that you can, with hard work and the

assistance of grace; but do it for yourself, not for Daddy or me.

To my beloved Matthew Christopher: thank you for your commitment to me, in good times and in bad. Thank you for constantly reminding me of the vision that we share. Thank you especially for your emotional and practical support in my final stages of dissertation writing. To be on this journey of life together is both hard and good; God only knows where we are headed.

Finally, I dedicate this dissertation to my father- and mother-in-law, Tom and Patti Zwier. If it weren't for you, this dissertation probably would not have been written. By all worldly standards, what you did was crazy. You bought a house in a city unfamiliar to you, moved away from your home and friends, and devoted three years of your lives to caring for your grandchildren and putting up with their parents. Thank you for taking that leap of faith on us; we are forever in your debt.

1.0 INTRODUCTION

> *...it seems to me that one can make a strong case for the thesis that causation in the natural sciences (better: causation in nature) is primarily and on the whole of the manipulative type. ... The laws themselves from which the observed uniformity has been deduced are established by laboratory procedures—and not just by passive observations. Our knowledge of causes and effects in remote regions in space, or, as in geological or palaeontological research, in time is based on and 'mediated' by our knowledge of natural laws for which we have sufficient experimental evidence from our laboratories.*
>
> von Wright ([1973] 1993, 120)

Causal statements and the concept of "cause" are an important—if not essential—part of our language and our everyday activities. Philosophers have worked for many centuries to elucidate the meaning and usage of our ordinary causal statements. Along the way, philosophers have also made many attempts to revise and normalize causal statements and concepts in order to make them philosophically "respectable". With the advent of modern science, and especially as philosophy and empirical science came to be separated into distinct disciplines, a new concern presented itself: that of addressing the *scientific* respectability of causal statements and concepts, and in particular, the respectability of causal statements and concepts with regard to our most advanced and mature science, physics.

Thus began (and thus continues) the project of seeking out and identifying a *physically respectable notion of causation*. The project, as commonly understood, is an effort to explicate exactly what physics tells us about the nature of causal relationships. It is often accompanied by either the implicit or explicit assumption that our everyday usage of causal concepts is embarrassing in some regard and in need of correction in order to be rendered physically respectable. And the project does not stop there. Even accounts of causation that are relatively well-respected within philosophy—accounts that provide explications and norms for causal utterances and well-conceived analysis of their empirical support—are often deemed problematic from the point of view of physics. Thus, philosophical accounts of causation too are candidates for correction (or, for the more radically inclined, outright dismissal).[1]

On my reading, those engaging in the effort to characterize a physically respectable notion of causation have arrived at something of a consensus in identifying two requirements of the notion of causation if it is to be physically respectable. These two requirements are described below:

(1) *Causation is a property of closed systems.* In our most fully-developed physical theories, we are able to identify a set of variables that constitutes a "complete" state description of a system. Given the complete description of the state of some system at time t_0, we can use dynamical equations to calculate the complete state of the system at any later point in time t_f, as long as the system is closed to all outside influence. A time evolution of this type—and only one of this type—is causal:

> When we talk of the principle of causality in physics, ... [we think] in terms of theories which allow (at least in principle) the calculation of the future state of the system under consideration from data specified at time t_0. No specific reference to "cause" or "effects" is needed, customary, or useful, but it is understood that all the phenomena (or variables) which can influence the system have been taken into account in the initial specification, *i.e.*, that the system is *closed* (Havas 1974, 24).[2]

This particular requirement of the physically-respectable view, therefore, is the following: the temporal evolution of a system deserves to be labeled "causal" only if: (a) the system is closed to external influence; (b) the state of the system at some point in time can be fully described (at least in principle);[3] and (c) there is a fully-worked out dynamics with

[1] To my knowledge, Peirce ([1898] 1992) and Russell (1912) were the earliest to explicitly express this view (and to do so in such a provocative way), but there were 18th and 19th century philosophers and scientists who were concerned with the project of characterizing a scientifically respectable notion of causation. Examples include Laplace ([1814] 1902), Mill ([1843] 1973), Whewell (1847), Herschel (1851), Bain (1870).

[2] Emphasis in the original. Peter Havas, a theoretical physicist, delivered these remarks in 1973 at an American Institute of Physics conference entirely devoted to the topic of causality and physics. For a sense of the topics and goals of the conference, see Rolnick (1974).

[3] Reichenbach ([1932] 1978) emphasizes the importance of complete state descriptions to the notion of causation. He describes an "inductive principle of causality": the idea is that, through repeated experimentation with the same type of system, we iteratively refine our state description by adding additional parameters (causes) when necessary, expecting at some point to arrive at a "complete" description of the system that

which we can (at least in principle) calculate the temporal evolution of the system from the fully-described state.[4]

(2) *Causal evolutions display spatio-temporal continuity and obey relativistic limitations.* For a temporal evolution satisfying requirement (1) to be considered causal we also require that the evolution exhibit spatio-temporal continuity in time. Russell described this requirement in the following way:

> Events can be arranged in a four-dimensional order such that, when so arranged, they are interconnected by causal laws which are approximately continuous, *i.e.*, events whose co-ordinates differ very little also differ very little. Or rather: given any event, there is a series of closely similar events, in which the time-co-ordinate varies continuously from rather less to rather more than that of the given event, and in which the space-co-ordinates vary continuously from those of the given event (Russell [1948] 2009, 285).[5]

Russell calls a series of events displaying such a continuity a "causal line" and insists that some kind of structure or quality "persists" throughout the entire evolution. He also calls the differential equations that connect the series of events "causal laws" (see Russell ([1948] 2009, 275–277)).[6] Spatio-temporal continuity effectively acts as a locality principle, preventing "action at a distance"—specifically, any relation of influence between spacelike separated events. Thus, the speed of light acts as an upper bound on the speed with which causal processes can propagate:

> The Minkowski light cone can, with complete propriety, be called "the cone of causal relevance," and the entire space-time structure of special relativity can be developed on the basis of causal concepts (Salmon 1984, 141).[7]

---

includes all relevant parameters and requires no further changes to accommodate future experimental data.

[4]When Laplace ([1814] 1902) gave his famous statement of deterministic causation, he was stating precisely this view, except that he assumed determinism and phrased his view in terms of the state of the entire universe (which, by definition, is a closed system). Note, however, that with only a slight modification to the requirement stated here, the evolution of indeterministic systems can also be considered causal: a relationship in which we can calculate the *probability distribution* over a set of possible states at time t_f from a full state description at time t_0 would be considered a perfectly valid causal relationship.

[5]Russell continues on to offer the following caveat: "This principle, apparently, does not hold for quantum transitions, but it holds for macroscopic events, and for all events (such as light-waves) where there is no matter" (Russell [1948] 2009, 285).

[6]Both Reichenbach (1956) and Salmon (1984) followed and expanded upon Russell's description of "causal lines". Dowe (1992, 2000), following Salmon but building also upon the ideas from Fair (1979), specified the requirement that the "persistent" quality in a causal process be a conserved quantity such as mass-energy, linear momentum, angular momentum, *etc.*

[7]References to relativistic limitations on the rate of causal propagation are pervasive in the literature on physical causation. Norton (2007) surveys a number of authors and reduces all of their assertions about a "principle of causality" to relativistic considerations. See also Havas (1974); Winnie (1977); Dowe (1992, 2000); Ney (2009).

Spatio-temporal continuity and relativistic limitations go hand-in-hand as restrictions on evolutions that can be properly called "causal" in the physically-respectable sense.

Thus stands the current consensus on a physically respectable notion of cause. I shall call the conjunction of requirements (1) and (2) the "Consensus View of Physical Causation" (CVPC for short). Interestingly, subscribers to CVPC come from various fields. Theoretical physicists who subscribe to the view include Hawking and Ellis (1973), Havas (1974), Stapp (1974), Bell ([1975] 2004), and Haag (1996).[8] Philosophers of science include Russell ([1948] 2009), Nagel (1961), Winnie (1977), Fair (1979), Torretti (1983), Salmon (1984), Redhead (1986), Dowe (1992, 2000), Lange (2002a), and Ney (2009). Even metaphysicians sometimes refer to something akin to CVPC as a means for bolstering their particular account of causation.

Somewhat ironically, although CVPC can be traced back to a desire to make the notion of causation align with physics, it is CVPC that has led some philosophers to be *skeptical* of any substantive notion of causation in physics. Notice that, inasmuch as requirements (1) and (2) manage to reduce the notion of causation entirely to the language of theoretical physics, the notion of causation becomes superfluous. Norton (2007) argues precisely this point; for him, the use of causal language or "principles" in modern physics (which he takes to be mostly identical to requirement (2)), rather than making any factual or substantive claims, is merely a trivial act of labeling.[9] Russell (1912) used requirement (1) to argue for triviality; for him, claims about causation are trivialized by the fact that, according to (1), adequate specification of a physical system is so detailed as to make it unlikely that a causal claim can generalize to more than one case. And if there is no generalization of cause-effect relationships, there is no need for causal language or notions at all.

CVPC has other problems. One problem is debate over known physical phenomena that violate special relativity and what such violations might imply about the status of relativistic assumptions about causation. There are some physicists and philosophers who hold the relativistic portion of CVPC (requirement (2)) in doubt because of unanswered theoretical

[8]Since the 1970's, much work in theoretical physics involves attempts at developing nonlocal theories, but there is no clear consensus as to whether such theories are understood to simply violate the above-described notion of causality or if they would in effect propose an altogether new concept of causality. Still, it appears that subluminal interactions are those which most physicists are comfortable labeling as "causal".

[9]Butterfield (2007) also leans toward this view.

problems regarding nonlocal quantum phenomena.[10] But more importantly for my purposes here, perhaps the central objection to CVPC thus far is the apparent irreconcilability between its claims and ordinary causal claims. In the process of "correcting" our folk notion of causation, the new *physical* notion of causation seems to have gone so far that the assertions to which it leads us are at odds with our folk causal assertions. There are two main problems in this regard.

First, our fundamental physical theories are time-symmetric, but causal relationships in our ordinary sense are not. In a closed system fully specified in terms of its state at a particular time t_0, the state of the system at any other time (past or future) can be calculated. Hence, there is no "privileged" direction of time in our fundamental physical theories.[11] This lack of time-direction in closed physical systems appears to be at odds with a natural causal interpretation in which the future states (rather than the past states) are dependent on the present state of the system.

Secondly, there is a mismatch between the kinds of events and objects picked out as causes and effects in ordinary causal claims and the physical states that can satisfy the role of causes and effects in CVPC. Causal relationships (at least those of the sort that seem to be important to humans) usually pick out dependencies between events or objects which, from the perspective of fundamental physics, are extremely small and localized. For example, it seems reasonable for a person who is lactose intolerant to say that her eating of a dairy product on a given morning caused her upset stomach later that day. From the perspective of fundamental physics, however, an event token such as "eating a dairy product" is a vanishingly small piece of a much wider state of affairs that could lead to an upset stomach several hours later. According to special relativity, anything in the past light cone of the space-time event that constitutes the upset stomach is causally relevant to its

[10]For example, Butterfield (2007) and Maudlin (2011) are outlying dissenters regarding requirement (2). Both question the idea that relativistic causality is endorsed by contemporary physics.

[11]See Russell (1912) and Havas (1974) on this well-known problem. Note that this comment is limited to *fundamental* physical theories, because the determination of the past state of a system is obviously not possible for all ways of describing states, nor for all theories of the time evolution of a dynamical system. For example, the dynamical evolution of a gas will be time reversible if the state of a gas is described in terms of a point in a $6N$-dimensional phase space giving the positions and velocities of each of its molecules. If the state is instead specified in terms of macroscopic thermodynamic variables, the state of the system at a future time can be determined only with high probability, and is far more uncertain for past times. See Nagel (1961, Ch. 10).

occurrence. If it is determination of the effect (or at least determination of the probability distribution over possible effects) that we desire in a causal relationship, it would seem that nothing less than an entire space-like slice of the past light cone will suffice as the cause of a given event. But this kind of cause description seems like an absurd violation of our common sense about causation. It would mean that occurrences on Mars or in some distant part of the globe would have to be included in the description of the cause of the upset stomach. Even more strangely, it would seem that a description of the cause as a space-like slice of the past light cone would make everything that it describes *equally* relevant to the effect's occurrence; someone switching on a light bulb on the opposite side of the earth would be just as much a part of the cause as is the person's eating of the dairy product.[12]

But there is another more basic problem that has, as of yet, passed under the radar. CVPC does not even do what it is designed to do—*i.e.*, it does not provide us with a notion of causation that is responsible to physics! It does not do justice to all of the ways (and perhaps not even the primary way) in which causal notions are operative in physics. In my view, there are two main ways in which it fails to be responsible to physics. First, CVPC fails to provide an epistemology for the causal knowledge that it claims physics to possess. It gives us no account of how the experimental methodology of physics grounds that knowledge, nor does it give an account of the specific features of that methodology that grant *causal* character to that knowledge. Second, CVPC ignores an extremely prominent area of physics—classical thermodynamics—for which its two requirements are hardly applicable. The requirements of CVPC are inspired by mechanics, where dynamical laws are well known and processes are at the forefront. In contrast, thermodynamics (despite its name) lacks fully-developed equations of motion, and static equilibrium states are the focus of its theoretical paradigm. Thus, a view of thermodynamics through the lens of CVPC would be a distorted view. And it would be wrong to accept an account of causation that is inapplicable to such an important, resilient, and widely-applicable subdiscipline of physics.

However, this dissertation is devoted to a positive proposal rather than criticism of CVPC. I propose that we take a fresh look at the interventionist account of causation and its applicability to physics. Unfortunately, interventionist causation has largely been dismissed

[12] This problem is discussed by Russell (1912), Field (2003), and Ney (2009).

as a serious candidate for such an application. The reasons for this perception are related to the reasons that give CVPC its *prima facie* attractiveness. The concept of a closed system (as per requirement (1) above) is the pride of physics. A closed and fully-specified system for which we have complete equations of motion is where physics is at its best; there, we taste what it might be like to be Laplace's demon. Closed systems are also, by definition, systems which allow for no external interaction, and hence are the worst possible candidates for interventionist analysis. The problem is that we tend to forget and/or neglect those areas of physics for which we do *not* have complete equations of motion or for which it *doesn't make sense* to consider entirely closed systems. In those areas, CVPC makes less sense and interventionism finds its home.

In this dissertation, I will argue that an interventionist analysis of physics—an analysis that acknowledges both theory and experiment as partners in the enterprise of physics—has great philosophical appeal. In particular, I will argue that the interventionist account succeeds where CVPC fails:

1. *The interventionist account of causation provides an epistemology of causal knowledge in physics that is rooted in experiment.* The interventionist notion of causation is intimately tied to experiment. However, that link is more well-established in the social and medical sciences, where experiments are often explicitly and purposely designed for causal testing. The question of whether or not interventionist causal notions are operative or important to experimentation in physics has not been discussed or explored. I will show that some of the early advocates for experimental method within physics were very much in line with interventionist thinking (chapter 2), and I will give a philosophical account of the in-principle connection between interventionist causation and any kind of experimentation (chapter 3). I will also give several examples of experiments from the early history of thermodynamics that were performed in a rich context of causal debate. Furthermore, using Boyle's experiments as an example, I will argue that, by closely examining the experimental processes that give rise to more abstract formulas and theories, we can identify interventionist causal relationships and explain their epistemic justification (chapter 4).
2. *The interventionist account of causation is well-suited to the patterns of reasoning that*

are intrinsic to thermodynamic theory. I will argue that the interventionist causal notions so evident in experiment also remain present in the ways in which we use and apply current thermodynamic theory. In fact, I will argue that thermodynamic theorizing does precisely what is required by the interventionist account: it chooses a system and draws boundaries around it, where those boundaries are of a nature such that there is enough isolation that the internal dynamics can be approximately described, but interactions of various types between the system and the external environment are also allowed and monitored. Thermodynamics is an area of physics which does not have a time-evolutionary framework as its focus; rather, its focus is on equilibrium states and, paradoxically, it treats time evolutions by reference to those static states. As a consequence, those philosophers who have CVPC in mind have seen little evidence of causal relationships when considering thermodynamics. But the causal relationships are there, and as I will argue, they are there in a much richer sense than in reference to mere propagation or time evolution; seeing the causal relationships requires the interventionist perspective. I will show how the interventionist account of causation reveals that thermodynamic theory has a substantive causal structure (chapter 5).

3. *The interventionist account offers the prospect of a unification of "physically respectable" causation and "folk" causation.* In this dissertation, my overarching argument is that the interventionist account of causation is much more "physically respectable" than philosophers have recognized. By calling the interventionist account "physically respectable", I mean to make the claim that the interventionist account *explicates important parts of scientific practice*. It explains our causal reasoning in the context of experiment, and it also explains and accounts for certain ways in which we use and apply theory to concrete problems. And there is a further philosophical benefit. As discussed above, there are several well-known problems for reconciling ordinary causal claims and physical causal claims of the CVPC type. The interventionist account of causation, in contrast, is well-known for the ease with which it accommodates everyday causal claims. Thus, inasmuch as we can accept the interventionist account of causation as being applicable and relevant in physics, there is hope for a unified notion of causation across physics and the everyday. I will discuss this and other possible implications of my thesis in chapter 6.

In arguing the above points, this dissertation employs a two-fold strategy, historical and philosophical. The historical facts that I discuss in this dissertation are not unknown, but I present certain events and discoveries anew in light of the interventionist philosophical perspective. In this respect, the historical portions of the dissertation are largely an exercise in using interventionist concepts as historiographical tools for understanding some features of the way that scientists approached their experiments and the substantial struggles of theorists in dealing with their subject matter. I believe that the interventionist perspective turns out to be revelatory of both the methodological patterns used by the experimentalists that I discuss and the conceptual struggles faced by the theorists that I discuss. In addition, the revelatory nature of the interventionist perspective is itself an argument for the importance of interventionist tools in understanding certain features of thermodynamic theorizing and analysis of experimental data in physics more generally.

The philosophical portions of the dissertation are impossible to segregate entirely from the historical portions, but are mainly constituted by chapter 3, the latter half of chapters 4 and 5, and chapter 6. Chapters 5 and 6 will be of greatest interest to the philosophically-minded reader. As I see it, the philosophical import of this dissertation is that it comprises a significant step toward characterizing a view in which the interventionist account of causation is considered to be "physically respectable" along the lines sketched above. In doing so, it opens up a promising possibility: causation may not mean something wholly different in physics than it does in other sciences, and it may not even mean something wholly different in physics than it does in everyday life.

2.0 EXPERIMENT AS A TEST OF CAUSAL CLAIMS: A HISTORY

> *. . . before the 17th century, appeals to experience were usually based on passive observation of ongoing systems rather than on observation of what happens after a system is deliberately changed. After the scientific revolution of the 17th century, the word **experiment** . . . came to connote taking a deliberate action followed by systematic observation of what occurred afterward. [. . .] Although passive observation reveals much about the world, active manipulation is required to discover some of the world's regularities and possibilities. [. . .] Experimental science came to be concerned with observing the effects of such manipulations.*
>
> Shadish et al. (2002, 2)

It appears to be a widely accepted view that there is something about experiment as a method of investigation that grants a special kind of access to knowledge about causes and effects.[1] This view is central to contemporary interventionist accounts of causation. The contemporary phase of interventionist/manipulationist accounts began in the mid-20th century and has received its fullest treatment recently by Woodward (2003b).[2] Woodward expresses the central idea of the interventionist account of causation in the following way:

> A manipulationist approach to causation explains the role of experimentation in causal inference in a straightforward way: experimentation is relevant to establishing causal claims because those claims consist in, or have immediate implications concerning, claims about

[1]This view is particularly prominent and explicit in the statistics and social science literature. For example: "[Experimentation]. . . has been one of science's most powerful methods for discovering descriptive causal relationships, and it has done so well in so many ways that its place in science is probably assured forever" (Shadish et al. 2002, 26); and: ". . . experimentation is such a powerful scientific and statistical tool and one that often introduces clarity into discussions of specific cases of causation. . . " (Holland 1986a, 946).

[2]For the purpose of consistency, I will primarily use the name "interventionist account" rather than "manipulationist account" in this dissertation. Woodward (2013) notes that the term "intervention" was introduced into the manipulationist literature as a way of denoting a manipulation with the appropriate "surgical" features. The term "manipulation" still seems to be widely used, however, and I have no problem with the term "manipulation" itself, if it is understood and defined in the same way as an intervention (see chapter 3).

> what would happen to effects under appropriate manipulations of their putative causes. In other words, the connection between causation and experimentation is built into the very content of causal claims (Woodward 2003b, 35).

Take, for example, a simple causal claim of the form "A causes B". The interventionist account of causation says that the very meaning of this statement is to be understood as equivalent to that of the claim that, under at least some specifiable circumstances, if an ideal experiment could be performed in which the value of A were varied, then the value of B would also vary as a result. A convenient corollary of the interventionist theory is that, because the very meaning of causal claims is to be interpreted in terms of ideal hypothetical experiments, a causal claim such as "A causes B" can also be empirically *tested* using a real experiment in which A is varied and B observed, as long as the experiment is carefully devised and properly carried out.

The interventionist account has been refined by many philosophers and is able to accommodate far more intricate causal claims and experimental methods than in the simple example just given. I postpone my discussion of those refinements until chapter 3; in the present chapter I wish to make a historical point about the interventionist account. The thesis that I wish to argue is that, while some may trace the history of interventionist accounts back to explicit expressions in Gasking (1955) or Collingwood ([1940] 1998), or perhaps event to Bridgman (1927), it has a much deeper and richer history. The idea of the link between experiment and causation is not a novel invention of contemporary philosophy of science; it has always been central to modern science, with roots stretching back to the beginnings of the scientific revolution. It seems to me that the historical background of the interventionist view is underappreciated (if even recognized at all) by critics of the interventionist account, and that even among proponents of interventionist causation, it has not been explored in detail.

The idea that experiment constitutes a privileged method for gaining knowledge of causes has a particularly rich and interesting history. Part of my aim in calling attention to that history is to argue against any temptation one might have to view interventionist accounts of causation just as a current philosophical fad, as marginalizable, or as just one of several equally valid candidates for fleshing out the meaning of causal claims. I will show that interventionist accounts are part of a long tradition of thinking about causation as empirically testable and intimately tied to experiment. More importantly, I will show that the tradition itself—*i.e.*, that of thinking about causation as tied to experiment—was by no means a sideline in the history of science; it is wrapped up in the fundamental ideas of the scientific revolution—the great reconceptualization of human inquiry into nature that resulted in what we call the New Science.

It is during a revolution that major thinkers articulate that which in later times becomes tacit. Thus, I begin this history by examining the thought of two of the main advocates for experiment during the early stages of the scientific revolution: Galileo and Bacon. I explore their explicit statements about the connection between experiment and knowledge of causes, and I also examine examples of experiments that they carried out and discussed in their writings. My examination will show that the turn toward experiment during the scientific revolution was marked by a sharp change in what was considered to be a valid cause. Galileo and Bacon worked on lessening emphasis on certain senses of causation prevalent in the academic Aristotelian philosophy of nature, while elevating a different sense of cause that was directly linked with the very experimental methodology that they were advocating. I then move forward in history and show that the narrowed experimental sense of cause that Galileo and Bacon promoted and characterized is strongly present in John Stuart Mill's famous list of experimental methods two centuries later.

Ultimately, my examination of these three figures in this historical thread of thought will reveal a great deal of constancy in the idea of the connection between experiment and causal knowledge. Along the way, I will point out both moments of conceptual progress and cases of unidentified methodological problems that will be left for later interventionists to resolve.

2.1 BACKGROUND: EXPERIENCE AS EVIDENCE

Experimental science—as a new mindset, a new set of practices, and a new formalization of the epistemic content offered by individually experienced historical events—was born during the seventeenth century.[3] Appeal to experience was not new in natural philosophy. Since ancient times, experience had been considered a source from which knowledge of the natural world was constructed, and in accordance with this tradition, the Aristotelian natural philosophy of the early modern period prided itself on its accord with commonsense experience that is perceivable by the senses. However, late sixteenth-century and early seventeenth-century Aristotelian appeals to experience were quite different from those that were to become common within the new, experimental natural philosophy that was prevalent at the end of the seventeenth century.[4]

Within early seventeenth-century Aristotelian philosophy of nature, nature was seen to have regular (but not exceptionless) modes of operation. Exceptions to the usual workings of nature were accepted as occasions on which the usual course of nature had been interfered with. Accordingly, experience was accepted as evidential support for natural philosophy inasmuch as it was a reference to *common* experience and "how things happen" in general. In contrast, singular occurrences—especially ones which appeared to be exceptional—had a much more dubious status as evidence, because there was no way to be sure that they represented the ordinary course of nature. In fact, for Aristotle, experience was not a name for individual perceptual events, but rather was formed of many similar instances of perception collected in memory: "Thus from perception there comes memory, as we call it, and from memory (when it occurs often in connection with the same item) experience [ἐμπειρία]; for memories which are many in number form a single experience" (*APo* B.19, 100a3-6; Barnes (1993, 73)).

Experiences which were contrived artificially or ones which used the aid of instruments were rare, but not unheard of; in fact, instruments were recognized as necessary in astronomy and optics. However, experiences involving the use of instruments, and especially those contrived using artificially constructed scenarios, did not have a straightforward epistemic place in the Aristotelian conception of science. Although some Aristotelians made attempts

[3] Crombie (1953) argues that the origins of experimental methods can be found in the medieval period. Crombie is right to examine and emphasize a continuity in scientific thought; however, it is undeniable that the seventeenth century marks a time of great change and genuinely novel patterns in the uses of experience in natural science. See, *e.g.*, Dear (1995, Ch.1).

[4] See Dear (1995) for an in-depth history of how the weight and types of experience accepted in natural philosophy changed over the course of the seventeenth century.

to incorporate such experiences into their epistemology of science,[5] the increasing prominence of such experiences in natural philosophical investigations were to become a challenge. A new epistemology, and new philosophical concepts to accompany it, were needed.

2.2 GALILEO GALILEI

Galileo Galilei is widely recognized as one of the major players in the development of the new experimental science. However, it is important not to overplay Galileo's role in the birth of experimentation, as if he overthrew the Aristotelian models of science and epistemology in one fell swoop.

First of all, there were extremely abstract and rationalist elements in Galileo's methods alongside his references to experience. Galileo did not have one consistent overarching methodology, nor was he much of a methodologist.[6] On the one hand, his thought was undoubtedly shaped by the Aristotelian ideal for science. In line with the Aristotelianism of Clavius and other Jesuits at the Collegio Romano, Galileo held up mathematical science as the model of certainty against which all other science (including natural science) must be measured.[7] Accordingly, he sought to found his new science of motion on true and evident principles. References to widely accepted items of common experience were ideal as support for the principles of his science of motion that he wished to establish, and thus, many of Galileo's references to experience can be read in line with the contemporary Aristotelian

[5]See Dear (1995, Ch. 2) on some of the distinctions proposed by Alhazen, Blancanus, Scheiner, and Aquilonius for handling individual, exceptional, and contrived experiences within an Aristotelian epistemology.

[6]It is well recognized that Galileo's methods and standards for scientific knowledge changed over the course of his career and were inconsistent even within the same work. See McMullin (1978, 219), who writes:

> There is . . . a temptation: to suppose that Galileo was in possession of a well-articulated coherent theory of science which he consistently employed throughout his entire scientific work. [. . .] [T]his is not the case. The methodological hints that he throws out ought to warn us against any such assumption. The same Salviati who, for example, praises Aristotle for having properly preferred sensible experience to natural reason is equally enthusiastic about Aristarchus and Copernicus for having made reason the mistress of their belief in defiance of the senses.

[7]See Wallace (1991, Ch. 1). For an in-depth examination of how Galileo's methods were in accord with an Aristotelian model of mathematical science, see Wisan (1978).

usage of common experience. He (and the characters in his dialogues) often appeal to generalities that should be part of common human knowledge on the basis of repeated sense experience; for example: "experience [*l'esperienza*] ... shows us that particles of fire ascend always by lines perpendicular to the surface of the terrestrial globe" (Galilei [1632] 2001, 41); we all have "the experience of seeing [*veder per esperienza*] that the speed of whirling has a property of extruding and discarding material adhering to the revolving frame" (Galilei [1632] 2001, 218); and "sensible experience [*la sensata esperienza*] shows that on earth there are continual generations, corruptions, alterations, etc., the like of which neither our senses nor the traditions or memories of our ancestors have ever detected in heaven" (Galilei [1632] 2001, 54).

On the other hand, there are strands of Galileo's thought that are undoubtedly novel and mark the birth of a new scientific approach to the world. Some of Galileo's references to experience broke from the Aristotelian mold only slightly. These experiences are like common experience in that they are accessible to anyone (with only moderate difficulty), but they may require more attentive or detailed observation than other items of common experience. For example, the experience of "a plummet hanging on a cord" confirms the principle that "the impetus acquired at any point in its motion is enough to carry it back to the height from which it started" (Galilei [1632] 2001, 25); and it is verifiable that "a glowing iron, which can surely be called hot, weighs the same and moves in the same manner as when it is cold" (Galilei [1632] 2001, 50). More striking, however, are the experiences that Galileo describes which involve a greater departure from standard Aristotelian experience: the intentional construction of artificial scenarios for the purpose of testing a claim. This type of experience is that which most resembles modern scientific experimentation, and it will be my focus here.

A couple of cautions are in order before I discuss Galileo's use of this last type of experience. First, we must be cautious in reading Galileo's language. Galileo's Italian did not have a word for what we now call "experiment"; he had only a single noun, *esperienza* (with its corresponding verb *(e)sperimentare*). There were indeed two separate words in Latin, *experientia* and *experimentum*, but a robust distinction cannot be read into them during the early seventeenth century, except in isolated authors.[8] A concept of experiment as distinct

[8]See Dear (1995, Ch. 1 & 2).

from other types of experience had not yet been formed, nor had its unique epistemic status been fully appreciated. Out of sensitivity to this point, I will refrain from using the word "experiment" in reference to Galileo's work; instead I will refer to "devised experiences" when he seems to be referring to the deliberate construction of an artificial scenario as an experiential test.[9]

A second caveat toward a nuanced understanding of Galileo's use of experience is the obscurity surrounding the relationship between his writings about experience and his actual practices.[10] Part of this has to do with the rhetorical style in which Galileo refers to devised experiences. His references to devised experiences in his published writings are rarely intended as historical accounts of individual occasions on which he performed an experiential test; rather, his appeals to these experiences seem to function more as claims about what the result of such a test would be, *were* it to be carried out. Such claims obviously reflect a great deal of confidence in the result, which may very well have come from Galileo's actual testing. However, they are not presented as such; in fact, the accounts are often construed more like thought experiments than actual experiments. We must keep in mind that, within the mindset of Aristotelian science, construing actual experiments as thought experiments was actually advantageous; it made the experience seem more evident on the basis of common experience and reason, and therefore provided the sense that the experience was unnecessary to test.[11] Galileo also de-emphasized novelty in the results of his devised experiences. He often wished to portray his results as something to be expected given well-known mechanical principles, as if everyone should have known all along what was to be expected. On some occasions—occasions which may seem curious to us—he actually seems to portray experiential test as unnecessary, given the certainty of the reasoning he has put forth.[12]

[9]Sometimes, Galileo uses the words "fare l'esperienza" [to *do/make* an experience] in referring to devised experience.

[10]That Galileo even performed devised experiences at all was put in doubt by Koyré ([1939] 1978). Since then, it has been well established through manuscript research (see, *e.g.*, Drake 1973) and the use of experimental history of science (see, *e.g.*, Palmieri 2009) that Galileo did indeed perform many of the devised experiences that he describes in his writings. In this paper, I will avoid the question of which devised experiences he actually carried out and which he did not. Instead I will focus on the logical structure of the experiences recounted and how they were intended to aid in the identification of causes.

[11]Similar points about the rhetoric of the thought experiment can be found in Naylor (1989, 124) and Dear (1995, 62).

[12]For example, in the midst of a long discussion of a possible devised experience in which a stone is dropped from the mast of a moving ship, Salviati says, "Without experiment [*senza esperienza*], I am sure

2.2.1 Galileo's devised experiences to determine a cause

The above cautions being noted, I will now pick out one strand of Galileo's thought: the idea that a carefully crafted devised experience can be a tool for the identification of causes, and can indeed be a test for causal claims.[13] I do not claim that this notion is a consistent guiding methodology in all of Galileo's scientific work and writings; however, it is recurrent enough to merit attention. Its greatest significance for my purposes here lies in its being the beginning of a long tradition of conceiving of causes and experiment as being interconnected.

The central feature of a devised experience is the intentional construction of a scenario—often using a mechanical artifact—to test or establish a claim.[14] As mentioned above, devised experience bears a much stronger similarity to modern scientific experimentation than either of the types of experience already discussed. More importantly, devised experience is intimately tied to Galileo's developing notion of a testable cause. Here I will analyze several examples of the devised experiences that Galileo discusses in his works and the causal conclusions that he draws from them.

---

that the effect will happen as I tell you, because it must happen that way; and I might add that you yourself also know that it cannot happen otherwise, no matter how you may pretend not to know it—or give that impression" (Galilei [1632] 2001, 168).

[13] My main point in examining this strand of Galileo's thought is very similar to one argued by Ducheyne (2006):

> It is my claim that Galileo was trying to construct a *new scientifically useful notion of causality*. This new notion of causality is an *interventionist notion*. According to such a notion, causal relations can be discovered by actively exploring and manipulating natural processes. In order to know nature, we have to intervene in nature. Generally: if we wish to explore whether A is a cause of B, we will need to establish whether deliberate and purposive variations in A result in changes in B. If changes in A produce changes in B, the causal relation is established (Ducheyne 2006, 443).

My analysis differs from Ducheyne's in two ways. First, I do not believe that Galileo was very consistent in his methodology for discovery of causes, nor did he seem to be altogether conscious of its novelty. Secondly, I prefer to call his notion of causation a *precursor* to the interventionist notions, since it is still relatively underdeveloped; there are significant methodological difficulties that he was unable to work through, as will become clear in section 2.2.2.2 below.

[14] Both Palmieri (2008) and Machamer (1998) stress the importance of the construction of artificial mechanical devices in Galileo's experimental philosophy. Palmieri (2008, 16) describes the strength of the artifact in empirical investigation in the following way: "The artefact has the power to create many similar circumstances for observing a phenomenon. [...] Patterns of phenomena can be physically realized at the observer's leisure by varying the control parameters. The artefact compresses the otherwise long-drawn-out, difficult formation of the observer's ability to predict about sparsely available phenomena." My emphasis here will be placed less on the experimental artifact itself, and more on the way in which Galileo uses such artifacts to display a contrast between two similar cases, such that a cause is made evident.

2.2.1.1 The cause of floating or sinking.

Galileo's *Discourse on Bodies in Water* was concerned with the question of whether it is the density or rather the shape of bodies that makes them float or sink in a fluid medium. Galileo's position was that the difference between the specific weight (*i.e.*, density) of a body and the specific weight of the medium is the only cause of the body's floating or sinking (see Drake 1981, 26). Galileo's former teacher, Francesco Buonamico, had an opposing theory. According to Buonamico, floating or sinking had two causes: (1) the resistance of the medium to division, and (2) the dominating element of the body in question, and its natural tendency toward or away from the earth's center (see Drake 1981, 60–61, 69–70). For example, on Buonamico's theory, an object made primarily of earth would have a greater tendency toward the center than water, and so it would tend to sink in water. But this effect could be tempered or counteracted by shape; water as a medium has a certain power to resist division, and this power of resistance responds differently to differently-shaped submersibles: a round or pointed surface would be able to penetrate the water more easily than would a broad, flat surface. Thus, according to Buonamico, shape plays a role in a body floating or sinking.

Galileo denied that shape plays such a role. One of Galileo's arguments that his cause was the true cause (and the only cause) involved a devised experiential test that would distinguish between the two theories.[15] He began by naming a methodological principle for differentiating between his causal claim and Buonamico's causal claim:

[15] It is interesting to note that in this case, experiential test was not Galileo's only argument for his causal theory, nor was it his primary argument. As has been noted by many scholars, Galileo demonstrated strongly rationalist tendencies throughout his career; in the *Discourse on Bodies in Water*, these tendencies can be seen in the geometrical demonstration he constructed based on several mechanical principles. In fact, he appears to have thought this demonstration to be sufficient to prove his causal claim (see Drake 1981, 59). Galileo also gave several rationalist arguments against Buonamico's theory to show why shape could not be a cause of floating or sinking. Interestingly, Shea (1977) suggests that the writing of the *Discourse on Bodies in Water* may have been an occasion which granted Galileo an increased sensitivity to the necessity of experiential checks:

> The errors that Galileo made in the first draft of *The Discourse on Floating Bodies* taught him that a mathematician can arrive at the right conclusion for the wrong reasons, and this heightened his consciousness of the regulative use of experiments. The printed version of the *Discourse on Floating Bodies* does not abandon the mathematical approach however: Galileo still moves from geometrised physical postulates by formal deductive steps to various theorems, but he avoids the exclusively abstract mathematical reasoning of Archimedies for the more inferential technique of introducing appeals to experience" (Shea 1977, 22).

> ...commencing to investigate with examination by exact experiment [*l'esame d'esquisita esperienza*] how true it is that shape does not at all affect the sinking or not sinking of the same solids, and having already demonstrated how a greater or less heaviness of the solid with respect to the heaviness of the medium is the cause of its ascending or descending, [then] whenever we want to make a test of what effect diversity of shape has on the latter, it will be necessary to make the experiment [*far l'esperienza*] with material in which variety of heaviness does not exist. For were we to make use of materials that could vary in specific weight from one to another, when we encountered variation in the fact of descent or ascent we would always remain with ambiguous reasoning as to whether the difference derived truly from shape alone, or also from different heaviness (Drake 1981, 74–75).

The idea behind the experience he recommends is the following: there must be a comparison of two cases that differ in only one respect, the shape of the submersible solid. If the two test cases have a different result, then it can only be attributed to the difference in shape, and it will be proven that shape is a cause of floating or sinking; if not, then it can be concluded that shape makes no difference to floating or sinking. Galileo next begins to give procedural instructions:

> Now wax is very suitable for this, since besides its receiving no sensible alteration from impregnation by water, it is tractable, and the same piece is very easily brought to any shape; while being very little less heavy than water, it can be brought to very nearly equal heaviness therewith by imbedding in it a few lead filings.
>
> Such material being prepared, make it for example into a ball the size of an orange or larger, and heavy enough to stay at the bottom, but so lightly that by the removal of a single grain of lead it will float, and then with that grain restored it will return to the bottom. Next, reduce that same wax to a thin broad leaf, and try the same experiment [*far la medesima esperienza*] again. You will see that, placed on the bottom with that added grain of lead, it remains there, and with that grain removed it will rise clear to the surface; but with the grain of lead restored again, it will sink to the bottom. And this same effect will always take place in all sorts of shapes, regular as well as irregular, nor will any ever be found that will come to float without removal of the grain of lead, or will sink to the bottom without its addition (Drake 1981, 75–76).

In the devised experience above, we have a quantity of wax with a few lead filings added to it, so that it its density nearly matches that of water. Then two trials are performed in which the wax is formed into a different shape (one a ball, and the other a thin, broad leaf). In both cases, the wax, which originally sits on the bottom, begins to float when one grain of lead is removed. It does this regardless of the shape into which it is formed, and so Galileo concludes that shape is not a cause of sinking or floating.[16]

[16] A shrewd methodologist will recognize that Galileo's conclusion here is not legitimate; this methodolog-

2.2.1.2 The cause of the increased strength of an armed lodestone. At a point late on the third day of his *Dialogo*, Galileo's characters discuss the fact that the connection between a lodestone and iron is greatly increased by encasing the ends of the lodesone in an "armature" of soft iron.[17] In discussing the cause of this increased strength, Salviati begins with a methodological principle:

> . . . since for a new effect there must be a new cause, we seek what is newly introduced by the act of supporting the iron via the armature, and no other change is to be found than a difference in contact. For where iron originally touched lodestone, now iron touches iron, and it is necessary to conclude that *the difference in these contacts causes the difference in the results* (Galilei [1632] 2001, 472, emphasis mine).

The "new effect" in the case at hand is a difference (specifically, an increase) in the strength of the connection between the loadstone and the iron object. Salviati's first move is to identify any other differences that could have led to this "new" difference. One possibility is that the addition of the armature increases the force of the lodestone, and the increased force is the cause of the stronger connection. But this is not the case, since armed lodestones show no difference in their attraction at a distance or the strength of their attraction at virtually the same distance.[18] Being unable to discern any other difference that could be responsible for the stronger connection, Salviati concludes that it is the difference in results comes "from the substance of the iron being finer, purer, and denser in its particles than is that of the lodestone", and that contact between two pieces of iron provides much more area of contact than contact between a bare lodestone and an iron object (Galilei [1632] 2001, 472). To test this causal claim, he recommends the following devised experience:

> What I am telling you (that is, that the great abundance of contacts made between iron and iron is the cause of so solid an attachment) is confirmed by an experiment [*esperienza*]. If we present the sharp point of a needle to the armature of a lodestone, it attaches itself

ical issue will be discussed more in section 2.2.2.2.

[17] This phenomenon was described by Gilbert in *De Magnete*, Ch. XVII. There, Gilbert describes a five-fold increase in the strength of the magnet. Galileo cites a more impressive result of an armed loadstone bearing eighty times more than it can bear without armature (see Galilei [1632] 2001, 470).

[18] See Galilei ([1632] 2001, 472): "In the first place, I am certain that the power and force of the stone is not increased at all by its having an armature, for it does not attract through a longer distance. Nor does it attract a piece of iron as strongly if a thin slip of paper is introduced between this and the armature; even if a piece of gold leaf is interposed, the bare lodestone will sustain more iron than the armature. Hence there is no change here in the force, but merely something new in its effect." Gilbert also showed that an armature makes no difference in the force of a magnet at a distance in *De Magnete*.

> no more strongly than it would to the bare lodestone; this can result only from the two contacts being equal, both being made at a single point. But now see what follows. A needle is placed upon the lodestone so that one of its ends sticks out somewhat beyond, and a nail is brought up to this. Instantly the needle will attach itself to it so firmly that upon the nail being drawn back, the needle can be suspended with one end attached to the lodestone and the other to the nail. Withdrawing the nail still farther, the needle will come loose from the lodestone if the needle's eye is attached to the nail and its point to the lodestone; but if the eye is toward the lodestone, the needle will remain attached to the lodestone upon withdrawing the nail. In my judgement, this is for no other reason than that the needle, being larger at the eye, makes contact in more places than it does at its very sharp point (Galilei [1632] 2001, 473–474).

Immediately after Salviati's description of the above "experience", Sagredo testifies, "I rank these experiments [*quest'esperienze*] with the needle very little lower than mathematical proof" (Galilei [1632] 2001, 474).[19] What entitles Galileo (writing in the voice of Sagredo) to such a high estimation of the value of this devised experience? The design of the "experience" is such that two different orientations of the needle are attempted: one with the pointed end of the needle touching the lodestone and the wider end touching an iron nail, and *vice versa*. The orientation of the needle is the only difference in the two cases; all else—the force of attraction of the lodestone, that of the nail, the time order in which the needle is suspended from each—is identical in the two cases. Thus, when a difference in the result is shown from these two different setups—*i.e.*, that the needle sticks to the object to which its larger end is attached—the difference that is proven to matter is the larger area of contact between iron and iron on the larger end of the needle.

2.2.1.3 The cause of cohesion of a solid. Early on in the *Discorsi*, Galileo's characters attempt to ascertain what it is that makes solid bodies resist breakage and remain in one piece. Salviati maintains that there are two causes of the cohesion of a solid body: repugnance to the void, which is in some sense a negative cause in that it produces its effect by being avoided; and a positive cause of "some sticky, viscous, or gluey substance that shall tenaciously connect the particles of which the body is composed" (Galilei [1638] 1989, 19). After Salviati proves satisfactorily to Sagredo and Simplicio through common experiences that a force due to the repugnance to the void does indeed exist, Sagredo asks why this force

[19]See Wisan (1978) for a discussion of the importance of mathematical demonstration as a standard for certainty in Galileo's work.

cannot, by itself, be enough to explain the cohesion of solids. Why, he asks, must there be another cause as well?

Salviati responds with a devised experience:

> I shall tell you first how to separate the force of the void from other [forces], and then how to measure it. To separate it, let us take some continuous material whose parts lack any resistance to separation other than that of the void. Water has been demonstrated at length, in a treatise of our Academician, to be such a material.[20] Thus when a cylinder of water is displaced [within a tube], and in drawing it, a resistance is felt against the detachment of its parts, no other cause can be recognized for this than repugnance to a void. In order to make the experiment, I have imagined an artifice which I can better explain by a diagram than by mere words. Consider $CABD$ here to be the profile of a cylinder of metal, or better of glass, empty within and very accurately turned, into the hollow of which there enters, with the smoothest contact, a wooden cylinder which can be driven up and down, of profile $EGFH$. This is drilled through the center so that through the hold there passes an iron wire, hooked at end K, while the other end, I, is broadened out in the shape of a conical screwhead. Things are so arranged that the upper part of the hole through the wood is indented in the form of a conical surface, shaped exactly to receive the conical extremity I of the iron IK when pulled down in the direction of K. Insert the wood, which we may call the piston EH, in the cylinder-hole AD, not so as to reach the upper surface of the cylinder, but to remain two or three inches away. This space is first filled with water, poured in while the vessel is held with its mouth CD upward, the piston EH then being replaced while the screwhead I is kept a little way from the indentation in the wood in order to allow the escape of air pressing against the piston, which will get out through the hole in the wood, this having been drilled a little larger than the stem of the iron IK. All the air having escaped, the wire is drawn back again, sealing the piston with its screwhead I, and the whole vessel is rotated to bring it with the mouth [CD] down.
>
> A container is now attached to the hook K, into which sand or some other heavy material is put, loading it until finally the upper surface EF of the piston is detached from the lower surface of the water, to which nothing held it joined except repugnance to the void. Then, by weighing the piston together with the iron, the container, and whatever it contains, we shall have the amount of the force of the void.
>
> Next, to a marble or glass cylinder of the same size as the cylinder of water we attach a weight which, together with the weight of the marble or glass itself, balances the weight of all the things weighed before. If this breaks the cylinder, we can unquestionably affirm that the void alone is cause enough to hold the parts of the marble or crystal together. But if it is not sufficient, and in order to break [the cylinder] we must add four times the above weight again, then we must say that the void offers one-fifth the resistance, while the other [resistance] is four times that of the void (Galilei [1638] 1989, 22–24).

The experience that Salviati describes is devised to test how much weight can be supported by a cylinder of water, and to compare this weight to the amount that can be

[20]Galileo is here referring to his *Discourse on Bodies in Water*, in which he argued that there is no resistance to separation in water, and that such a purported resistance could not be the cause of floating or sinking. See section 2.2.1.1 above.

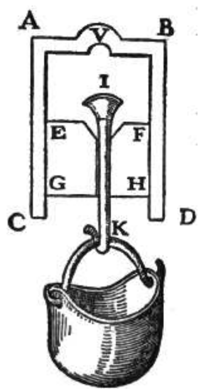


Figure 2.1: Galileo's illustration of a device for measuring the "force of the void". From Galilei ([1638] 1914, 14).

supported by another cylinder of the same size and shape but of a different material (*e.g.*, marble or glass). The important contrast in this experiment is between the case of the cylinder of water and the other cylinder (say, one of marble). Salviati begins with the assumption that water is a substance whose cohesion is due only to repugnance to the void, and not to any internal 'gluey substance'. Therefore, by measuring the weight that can be supported by a cylinder of water, one is measuring the force of the void alone. If it is found that the exact same maximum quantity of weight can be supported both by the cylinder of water and the cylinder of marble without breaking, then the experience proves that the strength of the cylinder of marble is due to the force of the void alone, and not to any additional cause. If, on the other hand, it is found that the cylinder of marble can support more weight than the cylinder of water, it is proven that there must be some additional force beyond the force of the void that accounts for the cohesion of the marble.

Galileo's reasoning, then, is the following: where there is a difference in the effect, there must be a difference in the cause. If there is no difference in the amount of weight that is supported by the two different cylinders, then there is no difference in the cause (*i.e.*, the force or forces that constitute the cohesion of a body). If, on the other hand, there is a difference in the amount of weight that is supported, there must be a difference in what causes the cohesion of the two cylinders. And since, on Galileo's assumption, water has only the force of the void holding it together, any additional weight that can be supported by a cylinder of marble must be due to another distinct cause.[21] Furthermore, both the force of the void and the force of this other distinct cause can be measured by the devised experiential test.

2.2.2 Galileo's notion of cause

In each of the above examples of a devised experience, Galileo's design relies on the notion of a cause as a *difference-maker*. In fact, Galileo gave exactly this definition of a cause in his *Discourse on Bodies in Water*: that "which, being present, the effect is there, and being removed, the effect is taken away" (Drake 1981, 130).[22]

The difference-maker conception of a cause is intimately tied to devised experience. Under this definition of cause, a devised experience is able to determine if something is a cause inasmuch as it is able to identify a difference maker; thus, the experience must be designed such that one thing is varied while keeping all else the same, and attention must be paid as to whether or not there is a difference in the effect. Take, for example, the following piece of causal reasoning from *The Assayer*:

If Sarsi wants me to believe with Suidas that the Babylonians cooked their eggs by whirling

[21] Later, when Salviati is questioned as to the nature of this distinct cause, he falters and gives a "fantasy" description of how this distinct force might actually reduce to tiny voids within the particles of the substance; the question is ultimately left unresolved. See Galilei ([1638] 1989, 26–28).

[22] Drake (1981, xxv) makes the claim that Galileo's 1612 definition marked the occasion in which "the word *cause* was first sharply defined for use in scientific inquiries". Galileo's definition of a cause as a difference-maker on this occasion is by no means an isolated idea in his works; he continued to reference it even in his works at the end of his career. For example, in the 1632 *Dialogo*, in discussing how a steelyard balance with unequal arms is able to support unequal weights, Salviati says, "Consider what there is that is new in the steelyard, and therein lies the cause of the new effect" (Galilei [1632] 2001, 249). In his discussion of the tides, Salviati cites the principle: "whenever a fixed and constant alteration is seen in the effect, there must be a fixed and constant variation in the cause" (Galilei [1632] 2001, 517). See also the above-cited principle that Galileo uses in his discussion of the armed lodestone: "since for a new effect there must be a new cause, we seek what is newly introduced by the act of supporting the iron via the armature, and no other change is to be found than a difference in contact" (Galilei [1632] 2001, 472). In the 1638 *Discorsi*, a difference-maker principle is operative in Galileo's discussion of the cause of different speeds in falling bodies: "If the difference in weight in moveables of different heaviness cannot cause the change [with distance] in the ratio of the speeds, because the heaviness does not change, then neither can the medium cause any alteration in the ratio of speeds since it too is always assumed to stay the same" (Galilei [1638] 1989, 77).

I am grateful to Paolo Palmieri for pointing out to me that Galileo may be merely repeating a common Latin saying here: "sublata causa tollitur effectus". Examples of this or a similar saying can be found in the works of Galen (*On Antecedent Causes*, VII, 76), Aquinas (*c.f. Super Sententiarum*, lib. 1 d. 38 q. 1 a. 5; *Summa Theologiae*, II–II, q. 20, a. 2), and Robert Burton (*Anatomy of Melancholy*, Part 1, Sect. 2, Memb. 1, Subsect. 1). Nonetheless, even if Galileo is merely citing an aphorism, he certainly gave the saying a new methodological significance using devised experiential tests.

> them in slings, I shall do so; but I must say that the cause of this effect was very different from what he suggests. To discover the true cause I reason as follows: "If we do not achieve an effect which others formerly achieved, then it must be that in our operations we lack something that produced their success. And if there is just one single thing we lack, then that alone can be the true cause. Now we do not lack eggs, nor slings, nor sturdy fellows to whirl them; yet our eggs do not cook, but merely cool down faster if they happen to be hot. And since nothing is lacking to us except being Babylonians, then being Babylonians is the cause of the hardening of eggs, and not friction of the air" (Galilei [1623] 1957, 272).

The above example obviously has a mocking tone, but the point is a simple and relevant one: something is a cause of a certain effect if there are two scenarios in which, all else being identical, (a) the something differs, and (b) the effect correspondingly differs. Therefore, an experience can be devised for the purpose of determining whether or not something is a cause of a given effect. In order to tell if an alleged cause C actually is a cause of a certain effect, we must set up scenarios that are identical except for C being present or absent. If the effect's presence or absence aligns with the presence and absence of the alleged cause, then we may deem it a genuine cause; if not, we may reject it as a cause.

2.2.2.1 Galileo's cause as an innovative advancement. Galileo's definition of a cause and his proposal that causal claims could be tested through devised experience was a novel advancement in natural science. The Aristotelian academics who were his contemporaries were certainly happy to speak of causes, and they were also happy to use examples from experience to bolster their philosophical claims. But Galileo's proposed method is significantly different: for him, devised experience constitutes a criterion of truth against which causal claims can be measured. A devised experience is more than a general reference to common experience, but rather a carefully crafted material test to see how various setups differ in their result. A cause, when tested in this way, is not a mystery hidden behind the veil of experience, but rather a difference made directly detectable to the senses. Devised experiences were, in Galileo's hands, an instrument allowing for the detection (and even measurement) of causes by the senses (see Drake 1981, xxxvii). Galileo's innovation was that of introducing the notion of an *empirically testable cause*.[23]

[23] Shea (1977, 39) writes, in agreement with my claim here, that, "It is one of Galileo's great contributions to the development of the scientific method that he clearly recognised the necessity of isolating the true cause by creating artificial conditions where one element is varied at a time."

In fact, the infamous passage from the *Discorsi* in which Galileo rejects of the causal investigation of acceleration is best interpreted not as a wholesale rejection of causal investigation, but rather as a rejection of the *style* of causal investigation in which his Aristotelian contemporaries participated:[24]

> The present does not seem to me to be an opportune time to enter into the investigation of the cause of the acceleration of natural motion, concerning which various philosophers have produced various opinions, some of them reducing this to approach to the center; others to the presence of successively less parts of the medium [remaining] to be divided; and others to a certain extrusion by the surrounding medium which, in rejoining itself behind the moveable, goes pressing and continually pushing it out. Such fantasies, and others like them, would have to be examined and resolved, with little gain. For the present, it suffices our Author that we understand him to want us to investigate and demonstrate some attributes of a motion so accelerated (whatever be the cause of its acceleration) ... (Galilei [1638] 1989, 158–159).

In the above passage, Galileo declines to participate in a style of discourse that he considers to be of "little gain". He never states that the question itself is worthless, or that causal investigation in general is a vain effort; his point is rather that the proposed causes of acceleration, which cannot be tested through devised experience, are not worthy of a dispute. Unless a causal claim is put in a form such that it can be tested through a devised experience, it is a mere "fantasy".[25] Another of Galileo's criticisms of an Aristotelian causal account is

[24] Drake (1981, xxviii) acknowledges that there is significant interest in the discussion and discovery of causes in Galileo's work up to and including the 1612 *Discourse on Bodies in Water*, but he argues that in Galileo's later works, his "analyses were less causal and less experimental than in his analysis in the *Discourse*". He calls it "Galileo's mature position—that causal inquiries might be well abandoned in physics" (Galilei [1638] 1989, xxxiv) and claims that "rejection of causal inquiries was Galileo's most revolutionary proposal in physics, inasmuch as the traditional goal of that science was the determination of causes" (Galilei [1638] 1989, 159, footnote 12; see also Drake 1975). Clavelin (1974, 383) argues for a similar position: that Galileo's major innovation involved his "conceptual system in which rational necessity took the place of physical causality". I disagree with Drake's and Clavelin's assessments. Galileo did not reject causal investigation or causal theorizing (as evidenced by the examples I have cited and many others beyond); rather, he rejected the traditional type of causal discourse in which Aristotelians engaged and proposed a new standard for causal theorizing—testability through devised experience. See McMullin (1978), Wallace (1972, Ch. 5), and Wallace (1991, Ch. 2) for further support of Galileo's interest in causation. Palmieri (2008, Ch. 1) describes three stages in Galileo's thought regarding experiments and causes, in which the second stage involves "direct investigation of causes with the help of artifacts", while the third and last stage is one of discovery of non-causal principles through experiments with artifacts; he does not, however, claim that Galileo rejected the search for causes in his later years.

[25] Salviati's criticism of other proposed causes of the tides are along these same lines:

> ... among all things so far adduced as *verae causae* there is not one which we can duplicate for ourselves by means of appropriate artificial devices. For neither by the light of the moon or sun, nor by temperate heat, nor by differences of depth can we ever make the water contained in a

of a similar vein. Recall that on Buonamico's theory of floating bodies, one of the causes of floating or sinking was the predominant element in the body in question, and its tendency toward (or away from) the center of the earth. Galileo criticizes this causal account in the following way:

> I say that it is the same thing to consider in a movable the dominant of [two or more] elements as it is to consider the excess or defect of heaviness in relation to the medium, since in such action the elements operate only to the extent that they are heavy or light. ... Indeed, the immediate cause is its being less heavy than water, and the predominance of air is the cause of less heaviness, so that whoever offers as the cause the predominance of this element adduces the cause of the cause, not the proximate and immediate cause. Now, who does not know that the true cause is the immediate cause, and not the mediate? (Drake 1981, 70)

Galileo appears to be arguing in this passage that his cause (the heaviness or lightness of the body in comparison to the medium) is the true cause because it is "immediate", while the predominance of a certain element is "mediate". What does "immediate" mean for Galileo here?

> ... he who alleges heaviness brings forth a cause well known to our senses, because he can very easily ascertain whether ebony, for example, or fir, is heavier or less heavy than water; but who will make manifest to us whether the element of earth, or that of air, has predominance in them? Certainly there is no better experience of this than to see whether they float or go to the bottom. So that whoever does not know that such a solid floats unless he [first] knows that air predominates in it, does not know that it floats until he sees it float. For he knows it floats when he knows air has predominance, but he does not know that air predominates except when he sees it float, and therefore he does not know that it floats except after having seen it float (Drake 1981, 72).

Immediacy, for Galileo, seems to refer to accessibility to detection or measurement by the senses. The heaviness of an object can be measured by using a scale, but the ratio of the four basic elements in that object—*i.e.*, the cause on Buonamico's theory—is can only be indirectly inferred from such a measurement. The ability to detect or measure a property with the senses makes that property a superior candidate for a cause; a factor that cannot be measured or detected, and which thus cannot be subjected to a devised experiential test

motionless vessel run to and fro, or rise and fall in but a single place. But if, by simply setting the vessel in motion, I can represent for you without any artifice at all precisely those changes which are perceived in the waters of the sea, why should you reject this cause and take refuge in miracles? (Galilei [1632] 2001, 489) is a less worthy (if not completely unworthy) proposal as a cause than a factor that can be subjected to such a test.[26]

It is interesting to note that the idea of a cause as a difference-maker does not necessarily imply abandonment of Aristotelian causes. Galileo was familiar and well-versed in the use of Aristotelian causes as part of natural philosophy (as he shows when he writes in Simplicios voice, particularly in the *Dialogo*), and he probably had no intention of dismissing Aristotelian causes entirely.[27] However, when he writes in the voice of Salviati, his use of the notion of cause is much more pragmatic: his concern is how we can *know* of a cause, how we can *test* it. Some Arisotelian causes might conform to the criterion of empirical testability, and some may not. In fact, many of Galileo's mathematical demonstrations seem to offer perfectly acceptable Aristotelian formal causal explanations, and these are testable through devised experience.[28]

For example, one major topic of the *Discorsi* is the scaling problem: why is it that, as an object is scaled up proportionally in all three dimensions, its strength does not scale in such a way that it can support a proportional amount of weight? Specifically, what causes a beam to break under its own weight when it reaches a certain size? Galileo explains that, assuming consistent density and strength of the material, the force of resistance that the object's cross section can offer increases as the squared ratio of a single dimension, while the weight of the object increases as the cubed ratio of that dimension (see Galilei [1638] 1989, 121–122). In giving this explanation, Galileo has given a perfectly acceptable Aristotelian formal cause, in the tradition of the mixed sciences. Yet the cause he has given is also perfectly acceptable on his own standards of testability. One can envision the following devised experience without departing much from Galileo's own text: Take two similar (*i.e.*, proportional) cylindrical beams AB and CD and suspend them each horizontally from one end, such that AB, which is bigger, is just barely big enough to break under its own weight. The difference in the

[26]For example, Galileo dismissed Kepler's proposal that the moon's attractive force was the cause of the tides because this was an occult cause. He criticizes the idea in his *Discourse on the Tides* (see Finocchiaro 1989, 128) and in the *Dialogo* (see Galilei [1632] 2001, 487).

[27]See Wallace (1991, Ch. 1) for an examination of the depth and extent of Galileo's Aristotelian influences. Wallace (1991, Ch. 2) deals with the various distinctions and maxims about causes that Galileo inherited from his Aristotelian education, and his own use of causal reasoning through *regressus*.

[28]Machamer (1978) argues that Galileo, working in the tradition of the mixed sciences, was concerned primarily with formal and final causes, and sometimes material causes.

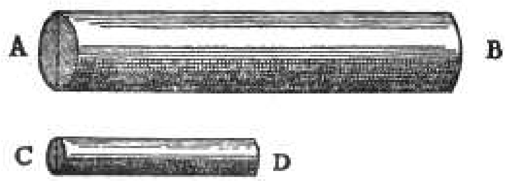


Figure 2.2: Galileo's illustration of two cylindrical beams of differing dimensions. From Galilei ([1638] 1914, 124).

effect (*i.e.*, breaking or not breaking) is due to the only other difference in the setup: that the ratio of volume to cross-sectional area is much greater in beam CD than it is in beam AB.

Other Aristotelian causes (including material and efficient causes) can also be ranked as legitimate Galilean causes to the extent that they can be shown to be difference-makers in a devised experience. To the extent that they make no claim of being difference-makers in a devised experience, they lack serious content and are mere fluff to be discarded along with notions like sympathy and antipathy (see Galilei [1632] 2001, 475–476). Hence, Galileo's notion of cause, while still allowing for some compatibility with Aristotelian causes, has the important advantage of being testable, and thus, is able to be established more rigorously.

2.2.2.2 The difficulties of innovation. Despite the advancement that Galileo offered in his new notion of cause and the promise that this innovation held for the "new science", there are several methodological issues related to the empirical testing of causes that he either never considered or was unable to resolve.

First, his writings betray an inconsistent attitude toward a 'one-cause-one-effect' principle. At several points in his works, he calls upon such a principle: "For any effect there is one unique and true and most potent cause" (Galilei [1638] 1989, 26).[29] For example, in his *Discourse on Bodies in Water*, Galileo was not interested in *a* cause of floating or sinking—*i.e.*, one of several factors that contribute to the phenomenon. Rather, he was

[29]See also: "...one effect can have only one basic cause" (Galilei [1632] 2001, 517); "...ultimately one single true and primary cause must hold good for effects which are similar in kind" (Galilei [1632] 2001, 485).

seeking *the* cause—*i.e.*, the single determinant—of floating or sinking. He stated his goal in the following way: "to explain what it is that is the *true, intrinsic, and entire cause* of the rising and floating of some solid bodies in water, or of sinking to the bottom" (Drake 1981, 25, emphasis mine). Nothing but a single and totally determining factor would satisfy.

In construing his causal search in this way, Galileo appears to presume that there would be a single, universal cause that would determine the effect of floating or sinking, no matter what the time, place, or other background conditions might be. He makes no mention of contingencies under which the sought-after cause might not produce its effect, demonstrating little sensitivity to the possibility that causal relationships may only hold within a limited range of circumstances. For example, in regard to the cause of floating or sinking, an exceptional circumstance that may immediately come to the mind of a modern is a dense object being supported on water by surface tension.

In the devised experience with wax described in section 2.2.1.1, Galileo concludes that shape cannot be a cause of floating or sinking because a difference in shape was tested in one case and shown not to make a difference to the result. For one who holds the 'one-cause-one-effect' principle, and who takes such a cause to be operative universally, this is a correct conclusion. But Galileo's analysis shows a neglect of the possibility that causes may operate only under a specific range of circumstances; considering this possibility, it cannot, in fact, be concluded from one case that something is not a cause if the circumstance tested is outside of the normal operation of that cause.

As it happened, this was precisely the argument and the counter-experience that Galileo's opponents used against his causal theory.[30] The following is Galileo's characterization of their argument:

> . . . it is necessary, if we wish to see what differences of shape may effect, first to select a material able by its nature to penetrate the bodily character of water; and for that purpose they thought a material would be suitable that, whenever reduced to spherical shape, went to the bottom; and they chose ebony. Making then a little chip of this as thin as a vetch pod, they showed how this rests when placed on the surface of water, without sinking to the bottom, while on the other hand they showed that a ball made of the same wood and no smaller than a hazelnut does not stay afloat, but sinks. From that experiment [*esperienza*] they thought they could freely conclude that in the flat chip breadth of shape was the cause

[30]The counter-experience was devised by Lodovico delle Colombe; other opponents in the debate included Papazzone, Arturo Pannochieschi de' Conti d'Elci, Giorgio Coresio, and Vincenzio di Grazia (see Shea 1977).

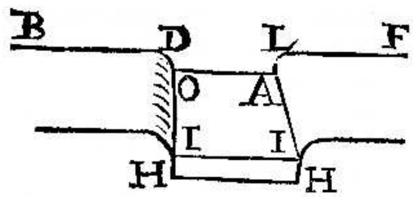


Figure 2.3: Galileo's illustration of why an ebony chip floats on the surface of water. From Galilei (1612, 34).

> of its not sinking to the bottom, inasmuch as a ball of the same material, differing from the chip only in shape, went to the bottom of the same water (Drake 1981, 80).

What Galileo's opponents had done is use his own methodological principle against him. If a devised experience is to be the test of a causal theory, they had only to find and propose a devised experience in which the shape of a piece of ebony *made a difference* to its floating or sinking. With this, the claim that shape is a cause of floating or sinking would be satisfactorily proven.

That Galileo felt threatened by the proposed experience is evident in the lengths to which he goes to explain why it does not succeed in proving what it claims to prove. His ultimate explanation is that the ebony chip is actually entirely below the surface of the water, so that the relevant density to consider is actually of that of the ebony plus the air above it, and that the ebony-plus-air complex floats because it is less dense than water. Galileo's solution, although *ad hoc*, is methodologically legitimate according to his own criteria for a devised experience: he argues that there is not just one difference (*i.e.*, shape) that could be producing the difference in effect—there is actually another, overlooked difference—that the ebony chip is floating within a canyon of water, so that its surface is actually below the water level. Thus, the effective density of the object is lowered by the added cushion of air below the surface of the water.

Instead of crafting this explanation, however, it might have been perfectly acceptable for Galileo to admit that there was a circumstance—*i.e.*, interactions at the surface of water—in which another cause was relevant to the floating or sinking of a solid body. But Galileo adamantly resisted this admission because of his commitment to his cause as the "true, intrinsic, and entire cause". In fact, he was so certain that his cause was the true and only one that he proceeded to calculate the heights of the 'ramparts' of water for various dense materials floating on the surface of water, without any empirical testing:

> Now that the true cause has been found for the floating of those bodies that otherwise, as heavier than water, should sink to the bottom, it appears to me that for a complete and distinct knowledge of this affair it will be good to proceed demonstratively, discovering those particular events that take place concerning these effects, investigating the ratios that bodies of different shapes and different materials must have to the heaviness of water in order that they may, by virtue of the contiguous air, remain afloat (Drake 1981, 123).

Galileo's extreme confidence in the 'one-cause-one-effect' principle gave him confidence in his ability to predict and calculate the results of other possible experiments—confident enough as to not feel any need for empirical testing.[31] In this instance, Galileo appears to have believed that once a cause is found, it is found; there is no need to search for the circumstances under which it operates (or does not operate) or the specific details of its operation. There appears to be no expectation that such details might be relevant at all!

It seems however, that he softened his view somewhat on the 'one-cause-one-effect' principle as his career advanced. In the *Dialogo*, he appears to contradict the principle during the discussion of the cause of the appearance of dark and light spots on the moon: "...because there are more ways known to us that could produce the same effect, and perhaps others that we do not know of, I shall not make bold to affirm one rather than another to exist

[31] Shea (1977, 27) reflects on this passage:

> It is manifest that Galileo is allowing himself a mathematical holiday. Having laid down a principle, he draws all the possible logical conclusions, untrammelled by any reference to the practical problems of physical application. This is not because experiments are difficult, tiresome or simply liable to erroneous interpretations. The real reason is that Galileo is convinced that once the cause of a particular phenomenon has been ascertained, the outcome of experiments can be predicted before they are performed.

And also:

> His faith in the uniformity of nature and his certitude that there is one basic cause implies a belief in the sufficiency of one or at the most of a few well-chosen experiments. [...] Galileo considers this sufficient to establish the relevance of geometrical reasoning and to allow him to proceed deductively. The outcome of further experiments can be safely predicted from his armchair: this explains why he insists on experimental verification and yet is so content with so few actual experiments (Shea 1977, 43).

on the moon" (Galilei [1632] 2001, 113–114).[32] Galileo's admission here seems to be that several possible causes could each be individually responsible for an identical effect.

Later, in his discussion of the cause of the tides on the fourth day of the *Dialogo*, Galileo seems to admit of the possibility that several causes could *together* be involved in producing the total effect. He introduces his account in the following way:

> . . . from those accounts which we are sure of, and which happen to cover the principal events, it seems to me possible to arrive at the true and primary causes. I do not presume to be able to adduce all the proper and sufficient causes of those effects which are new to me and which consequently I have had no chance to think about; [. . .] though in other seas remote from us events may take place which do not occur in our Mediterranean, nevertheless the reason and the cause which I shall produce will still be true, provided that it is verified and fully satisfied by the events which do take place in our sea; for ultimately one single true and primary cause must hold good for effects which are similar in kind (Galilei [1632] 2001, 485).

We can see in this quote that although he still emphasizes the idea of a "true and primary" cause, he no longer claims that it will be the only factor to account for all of the variation in the tides. There is a fundamental cause which enables the result to occur in the first place, and once it has its effect, other "concomitant" causes (*e.g.*, depth and shape of a sea) may produce variations in the effect. He seems to have a kind of modularity concept in mind, in which several causes can combine in order to produce a total effect:

> Now this [the acceleration and retardation of motion of the earth] is the most fundamental and effective cause of the tides, without which they would not take place. But the particular events observed at different times and places are many and varied; these must depend upon diverse concomitant causes, though all must have some connection with the fundamental cause (Galilei [1632] 2001, 497).

Further development of the concept of modularity, and development of methods for robust testing of several causes under a wide variety of circumstances, would become a task for later scientists.

[32] See also the end of the *Dialogo*, where Simplicio refers to the unfathomable causes that God could have used to produce the motion of the tides:

> I know that if asked whether God in His infinite power and wisdom could have conferred upon the watery element its observed reciprocating motion using some other means than moving its containing vessels, both of you would reply that He could have and that He would have known how to do this in many ways which are unthinkable to our minds (Galilei [1632] 2001, 538).

Another methodological issue that Galileo may not have reflected upon is the question of the validity of modeling and analogical reasoning in the investigation of causes. His theory of the cause of the tides is an example where this issue is evident. It would be practically impossible to devise an experience in which the motion of the earth is varied for the purpose of observing any resulting change in the tides; therefore, in place of such an experience, Galileo refers to small-scale models that simulate this type of test.

> ... [this] motion would occur when the vessel was moved without being tilted, advancing not uniformly but with a changing velocity, being sometimes accelerated and sometimes retarded. From this variation it would follow that the water (being contained within the vessel but not firmly adhering to it as do its solid parts) would because of its fluidity be almost separate and free, and not compelled to follow all the changes of its container. Thus the vessel being retarded, the water would retain a part of the impetus already received, so that it would run toward the forward end, where it would necessarily rise. On the other hand, when the vessel was speeded up, the water would retain a part of its slowness and would fall somewhat behind while becoming accustomed to the new impetus, remaining toward the back end, where it would rise somewhat (Galilei [1632] 2001, 493).[33]

The effect having been produced on the model, Galileo reasons by analogy:

> Now, gentlemen, what the barge does with regard to the water it contains, and what the water does with respect to the barge containing it, is precisely the same as what the Mediterranean basin does with regard to the water contained within it, and what the water contained does with respect to the Mediterranean basin, its container (Galilei [1632] 2001, 494).

Galileo considered this argument to be a very powerful one. Indeed, the modeling of a large and unwieldy system by a smaller and simpler one is a very powerful tool. But Galileo appears not to notice that it is vulnerable to important objections. As I have shown above, his method for investigating causes is based on the idea that something is a cause of a certain effect if one can create two scenarios which are identical except for a change in the purported cause, and the effect is also shown to differ. In the current argument, Galileo shows using a model that changes in the acceleration of a vessel have the effect of the water sloshing back and forth. It is a much more tenuous step, however, to argue from the conclusion about the model—*i.e.*, that changes in acceleration of a vessel cause "tides" of water inside

[33] Galileo later mentions (also in the voice of Salviati) that he has a "mechanical model" of the motion of the tides (see Galilei [1632] 2001, 500), but details about the model are not known. Palmieri (1998) attempts to reconstruct the experiments and models to which Galileo may have been referring.

the vessel—to the analogous conclusion about the earth and its tides. The very differences between the model system and the actual system (*e.g.*, differences in size, in directions and types of acceleration, atmospheric cushioning, *etc.*) are conditions upon which the purported cause-effect relationship could be argued not to apply. It turns out that causal investigations performed on models will be particularly vulnerable to this problem, and this issue too is left for future scientists and methodologists to work through.

2.3 FRANCIS BACON

Galileo's contemporary, Francis Bacon, is commonly taken to be another of the founding fathers of the experimental method. Unlike Galileo, Bacon's significance is not primarily that of an experimenter, but rather that of an influential writer consciously promoting a new methodology for philosophy of nature. Experiment was a central theme in Bacon's methodology; "experiment" in his recommended sense, had two advantages over other kinds of "experience".[34]

First and foremost, experiment is an "assistant to the senses":

> ...the subtlety of experiments is far greater than that of the senses themselves even when assisted by carefully designed instruments; we speak of experiments which have been devised and applied specifically for the question under investigation with skill and good technique. And therefore we do not rely very much on the immediate and proper perception of the senses, but we bring the matter to the point that the senses judge only of the experiment, the experiment judges of the thing. Hence we believe that we have made the senses (from which, if we prefer not to be insane we must derive everything in natural things) sacred high priests of nature and skilled interpreters of its oracles; while others merely seem to respect and honour the senses, we do so in actual fact (Bacon [1620] 2000, 18).

According to Bacon, the senses, on their own, are unreliable. Bacon's critique of the senses comes not from a Cartesian worry but rather from a methodological position that contrasts ordinary experience with a different kind of experience that grants true knowledge of causes and forms. In ordinary experience, a person is subjected to a series of combined and disorganized impressions. The person cannot help but form generalizations on the basis of

[34] In his writing, Bacon did carefully distinguish between the words *experientia* and *experimentum*.

these impressions, but inasmuch as they were presented to the person in a disorganized and inattentive fashion, any generalizations formed on their basis will not be trustworthy.

Thus, Bacon consciously criticizes the prevailing Aristotelian understanding of experience, in which experience is understood as common-sense generalizations about regular occurrences (see section 2.1). For Aristotle, the mind naturally collects individual instances of perception together in memory to form "experiences", and such experiences constitute a secure foundation on which to build a philosophy of nature. For Bacon, in contrast, the faculties of the human mind are prone to certain errors, and the mind must be cured and transformed and guided in order for it to be able to properly organize the sense perceptions it receives.[35]

Thus, Bacon sets experiment apart from ordinary experience in that it can be purposefully designed and organized around a particular question. Experiment is carried out with methodical intention. It is attentive and it purposefully organizes experience "specifically for the question under investigation".

> For by itself sense is weak and prone to error, nor do instruments for amplifying and sharpening the senses do very much. And yet every interpretation of nature which has a chance to be true is achieved by instances, and suitable and relevant experiments, in which sense only gives a judgement on the experiment, while the experiment gives a judgement on nature and the thing itself (Bacon [1620] 2000, 45).

When combined with experiment, the senses are given access to experiences specifically designed to cast light on a particular question.

The second advantage of experiment over ordinary experience lies in its ability to reveal nature's true operation:

> . . . we propose a natural history which does not so much amuse by the variety of its contents or give immediate profit from its experiments, as shed light on the discovery of causes and provide a first breast to feed philosophy. . . . we are making a history not only of nature free and unconstrained (when nature goes its own way and does its own work), such as a history of the bodies of heaven and the sky, of land and sea, of minerals, plants and animals; but much more of nature confined and harassed, when it is forced from its own condition by art

[35]Bacon's doctrine of idols is an attempt to catalog and diagnose various types of errors to which the human mind is vulnerable. Malherbe (1996, 79–80) sees Bacon's idols simply as those untested generalities at which the mind arrives when allowed to hasten to an "anticipation" about empirical data. According to Gaukroger (2001, Ch. 4), by understanding the idols it is possible to transform the mind and so cultivate the "persona of the natural philosopher".

> and human agency, pressured and moulded. ... Moreover (to be plain) we put much more effort and many more resources into this part than into the other, and pay no attention to men's disgust or what they find attractive, since nature reveals herself more through the harassment of art than in her own proper freedom. ... we are certainly looking for a kind of experience which is far more subtle and simple than those which simply happen. For we bring and draw many things out of obscurity which no one would ever have thought to investigate if he were not following the sure and steady path to the discovery of causes (Bacon [1620] 2000, 20–21).

Bacon says here that nature reveals itself more when it is "confined and harassed", when it is "forced from its own condition by art and human agency, pressured and moulded". Here again he is explicitly contradicting the standard Aristotelian philosophy of nature, which held that there was a distinction between natural courses of events and "artificial" circumstances that are the result of human interference. According to an Aristotelian, a proper study of nature is performed by observing nature unfolding in all of its normal complexity, without any interferences. Art and artifact, in contrast, are occasions in which human efforts overwhelm and shape nature into something it is not. To interfere with nature in order to study it would be counterproductive; the actions taken by the researcher would obscure nature rather than reveal it.[36]

But for Bacon, art and contrivance is the way in which we discover nature's true capacities which remain hidden when nature is allowed to take its own course.[37] Nature contains occult powers and unrealized combinations which can only be brought to light by art (*i.e.*, by man's devising).[38] And although it is human interference that occasions novel wonders, these wonders are still entirely the product of nature and are not in themselves artificial.[39]

In emphasizing the "harassment" of nature, Bacon draws a distinction which will later be seen as extremely important in interventionist accounts of causation. There are those

[36] The Aristotelian art-nature distinction is widely discussed and Bacon is generally seen as overturning it; see McMullin (1965); Rossi (1968); Broadie (1982); Dear (1995); Pérez-Ramos (1996). Newman (2004), however, offers a dissenting analysis according to which intervention-type investigations of the natural world were not foreign to Aristotelian thought and were, moreover, commonplace within alchemical methodology. He argues that Bacon's thought is understandable as part of the alchemical tradition.

[37] Weeks (2007) gives a thorough analysis of the relationship between art and nature in Bacon's thought, particularly with respect to the concept of "nature bound".

[38] "For in artificial things nature accepts the yoke from the empire of man; for these things would never have been done without man. A completely new face is given to bodies by human effort and agency, a different universe of things, a different theatre" (Bacon [1620] 2000, 224).

[39] "All man can do to achieve results is to bring natural bodies together and take them apart; Nature does the rest internally" (Bacon [1620] 2000, 33).

experiences "which simply happen", and there are those experiences which we *make* happen by human agency. It is the latter type of experience that Bacon calls "experiment", and it is experiment that Bacon identifies as being particularly valuable in the discovery of causes. As I will explain in more detail in chapter 3, interventionist accounts will later come to call this the distinction between observation and intervention, and will similarly arrive at a connection between experiment and causation. But for now, let's look more closely at the method that Bacon recommends.

2.3.1 Bacon's method

Bacon's recommendation for the steady path to the discovery of causes was to compile a natural and experimental history. But this should not be done in any haphazard fashion; in order that "the mind may be able to act on them" (Bacon [1620] 2000, 109), the data pertaining to the phenomenon of interest must be organized in tables as follows:

- The *Table of existence and presence* is a collection of as many cases one can find in which the phenomenon of interest is present. In this table, it is advantageous for us to find as wide a variety of cases, as disparate as possible.
- The *Table of divergence*. Here, we take the items in the first table and try to think of cases that are as similar as possible to those, but in which the phenomenon is absent.
- The *Table of degrees* is a list of cases in which the phenomenon exists in degrees. We start with objects which have lesser degrees of the phenomenon in question, or cases in which the phenomenon can vary in the same subject; then we advance to cases in which the phenomenon is very strong or present to a high degree.[40]

The example that Bacon chooses to demonstrate is an inquiry into the form of heat. In his table of presence, some of the instances he includes are simple observations related to heat such as flames, boiling liquids, and the sun's rays. The corresponding instances in the table of divergence are phenomena that are as similar as possible but with little or less

[40] These tables are described in Book II of the *Novum Organum*. See especially Bacon ([1620] 2000, 109–135).

heat: flame-like appearances such as *ignis fatuus* and St. Elmo's Fire,[41] cool liquids, and the moon's rays.

These tables provide a way of organizing experience according to a causal question. Although many of the instances are naturally-occurring instances of ordinary experience, once they are properly organized into a table, the mind can work on them due to the powerful contrasts made evident in the tables.[42] When we place Bacon's table of presence and his table of divergence side-by-side, what we have is basically a series of Galilean-type contrasts: each pair of instances are designed to be as alike as possible except for the phenomenon of interest. Equipped with the set of pairs, we can ask: what has changed across each case of contrasts, such that the heat which was present is absent?

But of course, Bacon does not limit his tables to observational instances alone. Some of the instances in his tables are *experimental* instances. For example, when he mentions the sun's rays can be concentrated to produce great heat with a magnifying glass, he recommends that it be tested whether or not any heat can be produced with the moon's rays in like fashion:

> Carefully try an experiment whether by means of the strongest and best-made burning glasses the rays of the moon can be caught and combined to produce even the smallest degree of heat. If perhaps the degree of heat is too subtle and weak to be perceptible and observable to the touch we shall have to try the glasses which indicate the hot or cold constitution of the air.[43] Let the rays of the moon fall through the burning-glass and be cast on the top of a glass of this kind; and take note whether a depression of the water occurs due to heat (Bacon [1620] 2000, 113–114).

In proposing this experiment, Bacon sets up a contrast between the use of a burning glass with the sun's rays and its use with the moon's rays. He does this in order to test whether or not there might be some similarity between sunlight and moonlight, such that each can produce heat, even if their capacity for producing heat in this manner might differ by degree. Since the main difference between the two proposed instances is the difference between

[41] *Ignis fatuus* means "foolish fire", also known as "will-o'-the-wisp"; St. Elmo's Fire is the glow that sometimes appears above the mast of a ship during an electrical storm.

[42] Bacon is explicit about the analysis that should be applied to the tables, which is as follows (see Bacon [1620] 2000, 127): (1) In the *Table of presence*, for every case in which the phenomenon is present, reject as *non-causes* any accompanying features that are absent. (2) In the *Table of divergence*, for every case in which the phenomenon is absent, reject as *non-causes* any accompanying features that are present. (3) In the *Table of degrees*, reject as a *non-cause* any feature that increases as the phenomenon decreases, or decreases as the phenomenon increases.

[43] Here Bacon is referring to an early version of the thermometer.

sunlight and moonlight, any resulting difference in the effect (*i.e.*, the heat produced) can be attributed to whatever difference exists between sunlight and moonlight.

Here is another example:

> . . . quicklime sprinkled with water seems to generate heat either because of the concentration of heat previously dispersed (as we said above about stored herbs), or because the fiery spirit is irritated and angered by the water, and some kind of struggle and rejection of the contrary nature takes place. It will be readily apparent which of these it is if we use oil instead of water; for the oil will have the same effect as water in forming a union with the enclosed spirit, but not in irritating it. Wider experiment should also be made with the ashes and limes of different bodies as well as by dropping different liquids on them (Bacon [1620] 2000, 117).

He notes as one of his instances in his *Table of presence* that quicklime (which we know as calcium oxide, CaO) generates heat when sprinkled with water. He proposes two possible causal explanations for this: (1) that there is some kind of unitive interaction between water and the internal heat in quicklime. Under this theory, moistening quicklime triggers the release of this internal heat in the same way as happens in a large, moist compost pile that can get so hot that it bursts into flames. Or, (2) the interaction is one of opposition; heat is created by some kind of antipathy relation between the internal heat of quicklime and water. The proposed test is to sprinkle oil on quicklime instead of water. The idea seems to be that oil cannot have an antipathy relation with heat (rather, it is sympathetic with heat—for example, it burns), and thus the result with oil will reveal what is happening in the case of water. In contrasting the case of water sprinkled on quicklime with the case of oil sprinkled on quicklime, we will have either a positive or negative answer to the question of whether or not the difference between water and oil has any effect on the heat produced when put in contact with quicklime. And the answer that question will provide a hint as to *what it is about the water* that causes heat when it comes into contact with quicklime.

Let's look at one more example experiment that Bacon discusses among what he calls "crucial instances". The question is whether the cause of weight is an object's "own nature", or rather the attraction of the physical mass of the earth.

> Take one of those clocks which move by lead weights, and one of those which move by a compressed iron spring; and let them be accurately tested, so that neither of them is faster or slower than the other; then let the clock that moves by weights be placed on top of a very high church, the other kept below; and let it be noted whether the higher clock moves

> more slowly than it did because the weights have less power. Let the same experiment be done at the bottom of mines deep below the earth, to see whether a clock of this kind does not move faster than it did, because the weights have increased power. If it is found that the power of weights decreases at a height and increases under the earth, attraction from the physical mass of the earth may be taken as the cause of weight (Bacon [1620] 2000, 163–164).

What Bacon is describing here is, again, a deliberate construction of contrasting instances. The two clocks are alike in all ways except for their height with respect to the mass of the earth. If the cause is the object's own nature, no difference will be seen in the time they keep. If the cause is an attraction by the physical mass of the earth, there will be a difference in the time they keep.

As can be seen from the examples discussed here, Bacon considered both observational and experimental "instances" to be suitable for his tables. And Bacon is quite explicit about *both* observations of "nature unbound" and experiments ("nature bound") belonging to natural history.[44] In fact, he sometimes refers to "natural and experimental history" rather than simply "natural history". On Bacon's view, we can have some knowledge of nature when it is left free to work as it pleases (assuming its instances have been appropriately organized), but we can learn far more when nature is "bound" by intervention.

2.3.2 Baconian methodology: Innovations and limitations

As we have seen, Bacon's proposed method has a certain affinity to Galileo's notion of cause; just as Galileo's notion of cause relies on a contrast between two cases in which something is present and then is taken away, Bacon's entire method is built on an assemblage of such contrasts. In this section, I will would like to point out some of the strengths and weaknesses of Bacon's method as used for the identification of causes.

Something akin to Galileo's basic method—that of comparing two identical situations, varying one thing, and seeing if the effect also varies—is present within Bacon's methodology, although within a more complex inference structure. It is quite evident in his citation of a phenomenon in many instances and contexts that Bacon is much more of an empiricist

[44] See "Preparation for a Natural and Experimental History" in Bacon ([1620] 2000) and also the quotations from Bacon on natural history given above in section 2.3. See also Newman 2004, Ch. 5 for a discussion of Bacon's criticisms of the existing natural history tradition and his proposed reforms.

than Galileo. Where Galileo satisfies himself with a single experiment plus a rationalist argument, Bacon requires example after example. Even in the context of a single example, the one about quicklime and water, he calls for wide experiment across a whole range of bodies and liquids. With such requirements, his method displays an increased sensitivity to background conditions; it won't fall prey to Galileo's error of assuming that a causal relationship holds in all instances.

Interestingly, Bacon doesn't hesitate to use heavily Aristotelian language; for example, he frequently talks about natures and forms, and relations of harmony versus relations of opposition. This is the kind of language that Galileo mocks at times and that Boyle, only a few decades later, will reject as completely nonsensical. Still, even in the context of this language, he is raising the bar on intelligibility by subjecting such terms to his detailed methodological analysis of contrasting instances.

It is also of note that Bacon's aim for the discovery of causes is to gain deep knowledge of phenomena and to empower man with the ability to put that knowledge to use:

> ...our way and method (as we have often said clearly, and are happy to say again) is not to draw results from results or experiments from experiments (as the empirics do), but (as true Interpreters of Nature) from both results and experiments to draw causes and axioms, and from causes and axioms in turn to draw new results and experiments (Bacon [1620] 2000, 90).

Bacon contrasts his aims with the traditional enterprise of logicians who work with theoretical propositions and syllogisms. He also contrasts his goals with that of "empirics", who seek only practical utility, and who prefer profit over illumination. Bacon's goal in defining his methodology is to attain a knowledge that gets at the heart of things: one that is both illuminating and practically useful. After all, for Bacon, true knowledge of the world cannot help but be both of these things (see Pérez-Ramos 1996).

Bacon's causes and hoped-for causal explanations concern the "forms" which his method is designed to identify. Forms play a causal role in that "human power cannot be freed and liberated from the common course of nature, and opened up and raised to new effectiveness and new ways of operating, except by the uncovering and discovery of such forms" (Bacon [1620] 2000, 128). But these forms may not be what we expect: for Bacon, they are the "ultimate ingredients" of nature, the basic material structure of a thing, the way in which its constituent material parts are disposed (see Gaukroger 2001, 140). Forms are things like heat and weightiness and color. He is explicitly *not* interested in "abstract forms and ideas, which are not defined in matter or poorly defined" (Bacon [1620] 2000, 128).[45]

Bacon's concrete and material understanding of forms as causes places limitations on his notion of causation and his recommended method for discovery of causes. Since causal explanations are founded on forms, where forms are understood as basic properties like heat and weightiness and color, it seems that for Bacon causal explanations are *not* to be expected for the behavior of more complex aggregates of these basic properties, or in macroscopic things. This expectation that causal relationships will hold only at a low or fundamental level stands in contrast to Galileo's expectation. Recall that Galileo had no such scruples in applying his notion of cause—he was willing to talk of causes in mechanics and was not really concerned with ascertaining microstructure or grounding his explanations in any kind of constitutionalist way.[46]

Based on the idea of these basic forms as causes, Bacon's method makes the assumption that an effect obtains *if and only if* the cause obtains. Bacon's causes are *both* necessary *and* sufficient for their effects. In his methodology, if there is a case where an effect is present but a purported cause is absent, this is grounds for dismissal of the purported cause. And if there is a case where the cause is present and the effect is absent, this is grounds for dismissal of the cause as well.[47] As we will see in chapter 3, this if-and-only-if principle will not be upheld by the interventionist account of causation.

Another weakness in Bacon's method and understanding of causation (at least on the historical thread that I am drawing here) is that Bacon does not seem to acknowledge the epistemic difference between observed instances that stand in a contrasting relationship and *artificial* instances that stand in a contrasting relationship. He seems to treat the observed instances and the experimental instances as equal. This realization may seem surprising

[45] Also: "the form of a thing is in each and every one of the instances in which the thing itself is; otherwise it would not be a form: and therefore there can be absolutely no contradictory instance" (Bacon [1620] 2000, 131).

[46] Although the seeds of interventionist reasoning are certainly present in Bacon's method, his foundationalist understanding of causes is not necessary to the interventionist account of causation and is possibly even at odds with interventionism. See chapter 6 for a sketch of my anti-foundationalist understanding of interventionist causation.

[47] Refer to footnote 42 above to see how his method enforces this condition.

given his emphasis on the importance of experiment over ordinary experience (see section 2.3). But as it turns out, for Bacon, experiment (understood as an act of art, a *binding* of nature) is of importance primarily for its ability to generate novelty—*i.e.*, to realize the latent powers of nature that would not otherwise be realized. He seems to simply think that intervention makes more phenomena accessible to us than would otherwise be.

As will become clear in chapter 3, the interventionist account of causation will see much greater value in experiment than simply the novelty that it generates. If the contrasts that Bacon is able to draw in his tables are a powerful way of identifying causes, experiments devised for the purpose of creating the very types contrast in which we are interested should be considered all the more powerful. For the interventionist, experiments are superior to mere observation both because of the control we have over the situation and because they are an exercise of our ability to randomly and arbitrarily apply interventions in such a way as to be independent of various properties of interest.

2.4 JOHN STUART MILL

The third and final historical figure that I will discuss as a precursor to the interventionist account of causation is John Stuart Mill. Although he stands at a much later point in history than Galileo and Bacon, the prominence and influence of his characterization of scientific method and its relation to causation merits treatment here.[48] I will show that there are two distinct concepts of causation operating in Mill's thought. One of these is a precursor to an interventionist notion of causation and the other is an evolutionary notion that is a precursor to CVPC (see chapter 1). I will argue that the interventionist notion of causation is the one that is more consistent with the scientific methodology that he recommends.

[48]Mill's thought is sometimes criticized as lacking originality, and there are several philosophers who perhaps better deserve acknowledgement for some of his ideas. In particular, the rules now known as "Mill's methods" owe much to Herschel (see Robson's "Textual Introduction" in Mill [1843] 1973 and Cobb 2012). Mill also openly recognized that his methods were based on ideas that Bacon had already discovered; in fact, he saw himself as codifying Bacon's ideas. Still, Mill's characterization of scientific method is useful here as an influential articulation of a prevalent view of the relationship between causation and experiment in the mid-nineteenth century.

2.4.1 Mill's two distinct concepts of causation

Mill gives his famous "four methods of experimental inquiry" as just a small piece of his larger *System of Logic*. Mill defines logic as "the science of the operations of the understanding which are subservient to the observation of evidence" (Mill [1843] 1973, 12). The inclusion of questions of evidence in this definition reflects Mill's conviction that logic, as an art and a science, should not be limited to mere inference from assertion to assertion, but must also concern itself with the "pursuit of truth". Thus Mill's system of logic covers both deduction and induction. Deductive logic, for Mill, concerns the "import" of propositions and inference of one proposition from other propositions. But as an empiricist, Mill holds that all inference ultimately redirects back to induction (Mill [1843] 1973, 283); all knowledge comes from experience. So inductive logic is for Mill the far more important concern of logic, as it concerns the evidential legitimacy of all propositions.

Mill holds that all induction is grounded on the "axiom of the uniformity of the course of nature" (Mill [1843] 1973, 310). All laws of nature relate what Mill calls "parallel cases" for which "that which happens once, will, under a sufficient degree of similarity of circumstances, happen again, and not only again, but as often as the same circumstances recur" (Mill [1843] 1973, 306). But laws of nature can relate phenomena in two distinct ways. Some laws of nature describe uniformities which exist among synchronous phenomena, while others describe uniformities in phenomena over time as they precede and succeed one another (Mill [1843] 1973, 323–324).[49] Uniformities of succession constitute *one* of Mill's two distinct concepts of causation:

> The only notion of cause, which the theory of induction requires, is such a notion as can be gained from experience. The Law of Causation, the recognition of which is the main pillar of inductive science, is but the familiar truth, that invariability of succession is found by observation to obtain between every fact in nature and some other fact which has preceded it. . . (Mill [1843] 1973, 326–327).

He further says this:

> The cause, then, philosophically speaking, is the sum total of the conditions, positive and negative taken together; the whole of the contingencies of every description, which being realized, the consequent invariably follows (Mill [1843] 1973, 332).

[49]In contemporary language, this is what we would call the distinction between equations of state and dynamical equations.

Thus Mill's "philosophical" conception of cause is a deterministic, evolutionary one.[50] When he turns to the question of methods for discovering such causes, however, a very different picture emerges. Mill's four methods are specifically designed as inquiries into the cause of a given effect or the effect of a given cause. They are as follows (see Mill [1843] 1973, 388–403):

1. *Method of Agreement*. Collect many instances of the phenomenon that have nothing in common, save one circumstance. That one circumstance in common is the cause (or effect) of the phenomenon.
2. *Method of Difference*. Seek out two instances that are common in all ways except in regard to the presence or absence of the phenomenon. If a difference is seen in the phenomenon occurring or not occurring, then that one difference is the cause or the effect of the phenomenon.
3. *Method of Residues*. If unable to isolate the causal factor of interest (as required by the Method of Difference), but if the effect of the other present variables is known, subtract the known effects to find the effect of the cause in question.
4. *Method of Concomitant Variations*. When two phenomena are observed to covary, then one may be the cause of another (or they are "connected through some fact of causation").[51]

Although he calls all of the above "Methods of Experimental Inquiry", he explains that one of them in particular is more uniquely the method of artificial experiment, and that is the Method of Difference. In the Method of Difference, the two instances being compared are exactly the same in all ways except in the presence or absence of the phenomenon under investigation. And according to Mill, it is difficult to ensure this unless we ourselves have set up the testing scenario and are well acquainted with the situation.

> In the spontaneous operations of nature there is generally such complication and such obscurity, they are mostly either on so overwhelmingly large or on so inaccessibly minute a scale, we are so ignorant of a rest part of the facts which really take place, and even those

[50] Notice that this concept of cause bears a strong resemblance to requirement (1) of CVPC as defined in chapter 1.

[51] Mill's caveat here is due to his recognition that some cases of covariation can be due to a common cause of the two phenomena rather than one phenomenon being the cause of the other. See further discussion below.

> of which we are not ignorant are so multitudinous, and therefore so seldom exactly alike in any two cases, that a spontaneous experiment, of the kind required by the Method of Difference, is commonly not to be found. When, on the contrary, we obtain a phenomenon by artificial experiment, a pair of instances such as the method requires is obtained almost as a matter of course.... It is, in short (as M. Comte observes), the very nature of an experiment to introduce into the pre-existing state of circumstances a change perfectly definite. We choose a previous state of things with which we are well acquainted, so that no unforeseen alteration in that state is likely to pass unobserved; and in to this we introduce, as rapidly as possible, the phenomenon which we wish to study; so that in general we are entitled to feel complete assurance that the pre-existing state, and the state which we have produced, differ in nothing except the presence or absence of that phenomenon (Mill [1843] 1973, 392–393).

The Method of Difference requires a strict similarity between the two instances compared, and artificial experiment is usually best able to approximate this. So one advantage of experiment is the superior epistemic position we have in relation to the phenomenon under study.

But Mill says there is an additional and crucial advantage of experiment over observation. He says that we can never be sure that one thing is a cause of another unless we have succeeded in *producing* the phenomenon artificially by means of the purported cause. Mill's reason for saying this is that he acknowledges the possibility of two phenomena being correlated even when one does not cause the other. He realizes that some associations between phenomena are due to a common cause rather than to one phenomenon causing the other. The only way to reliably distinguish between such associations and true causal relationships is by experimental production of the effect by means of the cause.

> Until it has been shown by the actual production of the antecedent under known circumstances, and the occurrence thereupon of the consequent, that the antecedent was really the condition on which it depended; the uniformity of succession which was proved to exist between them might, for aught we knew, be (like the succession of day and night) not a case of causation at all; both antecedent and consequent might be successive stages of the effect of an ulterior cause. Observation, in short, without experiment (supposing no aid from deduction) can ascertain sequences and coexistences, but cannot prove causation (Mill [1843] 1973, 386).

So for Mill, purely observational methods are evidence for a causal relationship, but only artificial production by means of an experiment can prove causation. Here, Mill explicitly recognizes the most important asset of artificial experimentation. When we are passively observing a system, it is harder to discern whether or not one thing is really a cause of another. Covariation alone is not sufficient evidence to conclude that one thing causes another. So he emphasizes that must also be possible to produce the effect by means of the cause, if we are to be certain that one thing causes another. He identifies this as the true advantage of experiment.

This position—*i.e.*, that intervention is crucial to distinguishing between correlation and causation—is the core claim of the interventionist account of causation.[52] In a relationship between two phenomena, the property by which one phenomenon can be produced by intervening on the other is *the identifying feature* of a causal relationship as opposed to any other regularity or uniformity. This is the core of the interventionist account of causation, as will be discussed in more detail in chapter 3. In emphasizing this point, Mill is advocating the interventionist concept of causation. This is a second, distinct concept of causation.

2.4.2 Sorting out the two concepts

Mill's first notion of causation, which he calls the "philosophical" notion, is obviously a deterministic and evolutionary notion of cause. He was, of course, writing during the heyday of determinism and was, like many others, enchanted with the idea. But when he comes to the question of methods for discovering causes, his views are at odds with that supposed philosophical ideal. His methods correspond to a search for a cause that satisfies the following description:

1. The cause C covaries with the effect E. (Either the presence or absence of the cause coincides with the presence or absence of the effect, or some quantitative measure of the cause covaries with some quantitative measure of the effect.)
2. It is possible to produce variations in E by means of variations in C. This eliminates the possibility (left open by the first criterion) that E is the cause of C, or that both E

[52] Modern interventionist accounts are slightly more subtle than this, as will become clearer in chapter 3. One feature of interventions that gives them this distinguishing power is that interventions can break or modify certain causal relationships that exist in the natural system, but leave intact others for the purpose of study. But a truly subtle interventionist understands that it is not the intervening itself that gives knowledge, as if the action itself gives some special insight. Rather, it is the specific way in which an intervention variable is related to the system that is important (specifically, its probabilistic independence from other causes of the effect being tested). In fact, one remarkable result of statistical treatments of interventionism is that, in cases where a causal variable has the same relationship to a system as does an ideal intervention, and this causal relationship is known, passive observation can reveal exactly the same causal knowledge as intervention.

and C are effects of a common cause.

Note that his methods do not necessarily require that the cause determine the effect. His methods require only that the cause "covary" with the effect and that there be a productive relationship between cause and effect. As far as Mill's method is concerned, C can qualify as a cause of E even if C doesn't determine E. It is not clear how Mill's "philosophical" notion of causation corresponds to his proposed method or informs it, and it is not apparent that Mill recognizes the discrepancy between his strict philosophical definition of a cause and his method.

Mill does comment, however, on the fact that there is something of a double-way of talking about causes in his day:

> ... we see that each and every condition of the phenomenon may be taken in its turn, and, with equal propriety in common parlance, but with equal impropriety in scientific discourse, be spoken of as if it were the entire cause (Mill [1843] 1973).

What Mill does here is distinguish between a common sense of cause and a scientific or philosophical sense. Mill's position seems to be that C is a "common-parlance" cause of E if and only if:

1. C covaries with E in at least some situations.
2. It is possible to produce variations in E by means of variations in C in at least some situations.

While C is a "scientific" cause of E if and only if:

1. C covaries with E universally and unconditionally.
2. C determines E.

Mill's "common-parlance" cause is an interventionist notion, and his "scientific" or "philosophical" cause is something very similar to CVPC (see chapter 1).[53]

Although we can identify important strands of interventionism in Mill's thought, we can also criticize the inconsistencies in Mill's view of causation by analyzing his position from the

[53] Collingwood ([1940] 1998, 301) notices that Mill has a double standard for talking about causes and defends a position in which the two standards of causation are related. Collingwood's position is the best attempt of which I am aware to reconcile an interventionist concept of causation with something like CVPC; however, it still suffers from some of the problems that I identified in chapter 1 and cannot avail itself of some of the benefits of a purely interventionist account that I will discuss in chapter 6.

more fully-developed modern interventionist perspective. Mill's so-called "scientific" cause is simply a conflation of completeness of knowledge and causation. According to the modern interventionist view, if the criteria for C's being a "common-parlance" cause are satisfied, then C is no less a cause of E for being incomplete. And in addition, an interventionist would emphasize that C is no less "scientific" a cause for being incomplete and operative only under certain conditions. Put simply, completeness of knowledge is separable from the question of what is or is not a cause. To presume otherwise is precisely the error of advocating a view like CVPC.

2.5 CONCLUSION

This historical tour through the interventionist strands of thought in Galileo, Bacon, and Mill has shown that the core idea behind the contemporary interventionist account of causation is not a new idea imposed on science, but rather an idea of central importance dating back to the scientific revolution. As such, interventionist ideas should not be dismissed lightly as just one of many co-equal potential definitions of causation. At least since the early modern era, experiment has been understood as intimately tied to the testing of causal claims. The details of the relationship between experiment and causation have only become clearer with time. In the next chapter, I will give a fuller explanation of the interventionist account of causation and show how it can be applied to arbitrary experiments in physics.

3.0 CAUSAL INFERENCE FROM EXPERIMENT: AN EPISTEMOLOGY

> *Imagine that the slates of science and philosophy were wiped clean and that we had to construct our understanding of the world anew. As part of that reconstruction, would we reinvent the notion of a manipulable cause? We think so, largely because of the practical utility that dependable manipulanda have for our ability to survive and prosper. Would we reinvent the experiment as a method for investigating such causes? Again yes, because humans will always be trying to better know how well these manipulable causes work. Over time, they will refine how they conduct those experiments and so will again be drawn to problems of counterfactual inference, of cause preceding effect, of alternative explanations. . . . In the end, we would probably end up with the experiment or something very much like it.*
>
> Shadish et al. (2002, 31–32)

In chapter 2, I highlighted a consistent thread of ideas about the connection between experiment and causal knowledge across three famous figures in the history of science and and history of philosophy of science: Galileo, Bacon, and Mill. I showed that the idea of *what it is* to perform an experiment remained relatively robust across these thinkers: to perform an experiment was to create an artificially devised scenario in which some human intervention is involved, but certain other "natural" relationships are left intact for study. I also showed that, despite some superfluous assumptions these thinkers imposed on the concept of causation, the notion of cause associated with experimental testing remained central and relatively consistent over this history.

In this chapter, I will show that the contemporary interventionist account of causation can be very naturally seen as the realization of this particular line of thinking about the connection between experiment and causal knowledge. The basic idea behind the interven-

An abbreviated version of this chapter has been published as Zwier (2013).

tionist view of causation is consistent with the original Galilean idea, Bacon's methods, and Mill's methods. Furthermore, when the interventionist account is combined with the recent development of a formal language for describing causal systems and experimental interventions, it becomes possible to understand and describe more precisely how, why, and under what circumstances experiments are able to grant causal knowledge.

However, there is an important question that has not been approached within the interventionist framework: how are we to understand the relationship between causal knowledge and *real* experiments as they are carried out in the course of scientific investigation? The interventionist account is explicitly intended as a conceptual clarification of *what it is to be* a causal relationship, and as a result, it relies heavily on hypothetical experiments explicated in terms of ideal interventions. Given the emphasis on highly idealized experimental scenarios, it seems reasonable to ask how the account might be applied to real experiments in order to assess the extent to which a given *arbitrary* experiment grants knowledge about causes in the specific and concrete context of the experiment. In particular, it doesn't seem reasonable to expect that a scientist will always have an explicit causal question in mind when designing and performing an experiment; we might wonder whether or not such an experiment still affords causal inference independent of the experimenter's intentions.

This chapter attempts to extend the interventionist account and deal with such questions. I will begin with an overview of the interventionist account of causation and the conceptual resources it provides. I then build up a schema for understanding the logical, observational, and operational components of an actual experiment and how these combine to grant (or not grant) knowledge about causal relationships. I will argue (1) that the set of causal inferences afforded by an experiment is determined solely on the basis of contrasting case structures that I call "experimental series"; and (2) that the conditions that suffice for causal inference obtain quite commonly, even among "ordinary" experiments that are not explicitly designed for the testing of causal claims.

3.1 THE INTERVENTIONIST ACCOUNT OF CAUSATION

Let us begin with the central core of the interventionist account of causation. The interventionist account, in its most basic form, is intended as an account of the *meaning* of causal claims:

> . . . the connection between causation and manipulation is proposed as an interpretation of what causal claims mean (or what must be the case for them to be true). The idea is that meaningful causal claims must have an interpretation in terms of what would happen in hypothetical experiments and that if a causal claim is true, what is predicted to happen in the relevant hypothetical experiment must in fact be what would happen (Woodward 2003a, 95).

What is the "relevant hypothetical experiment" for a given causal claim? Roughly, the idea is the following: for a causal claim such as "A causes B", the hypothetical experiment under consideration is one in which the variable or factor A is manipulated or changed in some way, and any corresponding change (or non-change) in B is observed. According to the interventionist account of causation, consideration of such an experiment is logically embedded in the very content of the causal statement itself, such that evaluation of the truth or falsity of the causal claim will be tied to an evaluation of whether or not a change in B would be seen if the experiment were to be performed.

This translation of a causal claim into a claim about a hypothetical experiment does a few important things for us, philosophically. First, it provides an answer to what might be called the *asymmetry problem*—*i.e.*, the question of what grounds the distinction between cause and effect.[1] On the interventionist account, cause and effect are distinguished by the fact that an experimental intervention on the cause will (under some range of circumstances) correspond to changes in the effect, while an experimental intervention on the effect will not correspond to changes in the cause. Second, the translation into a claim about a hypothetical experiment provides us with a distinction between a genuine causal relationship and other types of correlative relationships (sometimes referred to as "mere correlations", or otherwise

[1] von Wright ([1973] 1993, 107) gives a description of the problem. On von Wright's view, the interventionist idea of causation is a solution to the problem. Hausman (1986, 143) also sees the interventionist view as a possible solution to the asymmetry problem, but ultimately he attempts to reduce that distinction between cause and effect to a distinction in the correlative links that causes produce among their effects but which effects do not produce among their causes.

as accidental, derivative, or conditional relationships). If we know that variables A and B are correlated, this fact underdetermines the various types of relationships that may exist between the two variables. Assuming that the correlation is not a spurious result of sample or selection bias, there are three different kinds of relationships that may hold (and which may also hold in combination): (i) A is a cause of B, (ii) B is a cause of A, and/or (iii) A and B share a common cause C (or set of common causes).[2] All three of these distinct kinds of causal relationships imply distinct sets of facts about the way in which the variables or factors in question would respond to interventions. Specifically, if only relationship (iii) holds, then no intervention on A will correspond to changes in B, nor will any intervention on B correspond to changes in A. In fact, one of the central goals of the interventionist account is to provide such a distinction:

> I take the guiding idea of a manipulability approach to causation to be that lying behind the distinction we make between causal relationships and mere correlations is a concern to distinguish between, on the one hand, a relationship between A and B that can be used to manipulate (in the sense that if it were possible to manipulate A, this would be a way of changing B) and, on the other hand, a correlation that would simply disappear when we attempt to manipulate B by manipulating A (Woodward 2003b, 33).[3]

A third philosophical asset of the interventionist account is that it provides a framework for clarifying the meaning of causal claims: clarification of a causal claim is to be done by carefully describing the hypothetical experiment relevant to the claim (see Woodward 2003b; Rubin 1986). Distinctions between different causal claims can be made on the basis of differences between the hypothetical experiments to which they necessarily refer. In addition, careful description of different possible hypothetical experiments associated with a causal claim allows for comparison or contrast of different interpretations of that claim. A clear and contentful causal claim will make reference to a clear hypothetical experiment; a vague causal claim, in contrast, will make no such reference, or will leave open a greater number and variety of experimental interpretations. In the extreme, a purportedly causal

[2]The idea of the various types of "causal connection" was first suggested by Reichenbach (1956, 28) and has been further refined since. Spirtes et al. (2000, Ch. 9) gives a discussion of the various distinct acyclic causal systems that generate a correlation between two variables and the ways in which experiments can be used to discern among them.

[3]Emphasis on this point is shared by most philosophers working on and within the contemporary manipulationist account; see, *e.g.*, Gasking 1955, 483; von Wright [1973] 1993, 116–117; Menzies and Price 1993, 191.

claim that cannot be clarified in terms of a hypothetical experiment would not be considered a causal claim at all. Thus, the interventionist account even gives, in a sense, a criterion of demarcation between causal claims and non-causal claims.[4,5]

The interventionist account provides a specification for the hypothetical experiment that is necessarily referenced by a contentful causal claim; in particular, it places certain constraints on the intervention that produces change in the purported cause variable, A, and it also has something to say about the range of background conditions under which a positive result (an observed change in the purported effect variable, B) must obtain for the causal claim to be considered true. However, before entering into these specifics, I would like to pause and consider the similarities between the simple sketch of the interventionist account just given and Galileo's, Bacon's, and Mill's understanding of causes and methods for their testing.

Recall from section 2.2.2 how similar Galileo's proposed idea of a cause is to that of the interventionist account: for Galileo, A was a cause of B if two scenarios could be devised in which, all else being identical, (a) A differed, and (b) B correspondingly differed. The "all else being identical" clause was extremely important for Galileo, in that the difference between the two devised scenarios had to be isolated (by experimenter intervention) so that *only* A differed, and nothing else; otherwise, it would not be possible to definitively conclude that A was the cause of B, rather than something else. We also saw in section 2.2.2.1 that, in agreement with the contemporary interventionist account, Galileo seemed to regard as useless causal claims that were not put in a form such that they could be tested through devised experience, and called them mere "fantasies". Bacon's method, as we saw in section 2.3.1, expanded on an idea similar to Galileo's proposed test of causal claims; the basic method—*i.e.*, of comparing two identical situations, varying one thing, and seeing if the effect also

[4] Holland (1986a, 959) recognizes this point as an asset of his own broadly interventionist account when he says, "The experimental model eliminates many things from being causes, and this is probably very good, since it gives more specificity to the meaning of the word *cause*."

[5] By "causal claim", I mean to include both negative and positive claims about causation. The claim "A is *not* a cause of B" is just as much a causal claim as "C is a cause of D". Both claims are causal because both make implicit reference to the result (or null result) of a hypothetical experiment in which the cause variable in question (*i.e.*, A or C) is manipulated and the effect variable (*i.e.*, B or D) is observed. A non-causal claim, in contrast, makes no reference to a hypothetical experiment, or at least not the right kind of experiment. For details on the criteria for the "right kind" of experiment, see sections 3.1.1 and 3.1.2.

varies—was still present, although within a more complex inference structure. His *Table of existence and presence* and *Table of divergence*, when placed side by side in comparison, constitute a series of Galilean-type experiments that are, again, in line with the contemporary interventionist account of causation. Finally, as we saw in section 2.4.1, Mill was able to explicitly recognize the most important asset of artificial experimentation. Mill noted that when we passively observe a system, it is harder to discern whether or not one thing is really a cause of another. He recognized that covariation alone is not sufficient evidence to conclude that one thing causes another, since the two could be effects of a common cause. So, in complete agreement with the contemporary interventionist account, he emphasized the possibility of producing of the effect by means of the cause as the distinguishing feature of the causal relationship.

3.1.1 Defining cause in terms of hypothetical experiment

Now, looking beyond the similarities between the views of these historical figures and the basic idea behind the interventionist account, we are ready to discuss how the interventionist account offers a much more sophisticated and detailed way of describing the hypothetical experiments that are relevant to the evaluation of causal claims. First, I will state more precisely the criterion for X being a cause of Y, on the interventionist account of causation:

> Interventionist Cause: X is a *cause* of Y iff, under some set of background conditions $\mathbf{BC} = \{BC_1, BC_2, \ldots, BC_n\}$ having values $\{bc_1, bc_2, \ldots, bc_n\}$, given some set $\mathbf{S} = \{S_1, S_2, \ldots, S_m\}$ of variables other than X and Y that are held fixed at predetermined values $\{s_1, s_2, \ldots, s_m\}$, there is some intervention I on X that would change the value of Y.[6]

[6] "Cause", as I use it here and throughout this dissertation, corresponds to Woodward's "type-level contributing cause". The criterion itself that I give here is a modified and simplified version of Woodward's **M** (see Woodward (2003b, 59)). Woodward's **M** requires detailed knowledge of the path from X to Y and which variables are on and off that path; in the context of my discussion here, I do not wish to assume that an evaluator of causal claims always has that knowledge, and so I give a criterion that does not require it. In addition, my criterion is intended to be more faithful to the implicit criterion for causation in the mind of an actual experimenter who is—implicitly or explicitly—testing a causal claim. It seems to me that such an experimenter will very rarely will be thinking in terms of explicit causal models with causal chains and paths. More likely, the experimenter will be thinking in terms of a distinction between four sets of variables: (1) variables constituting the background conditions under which the experiment is carried out (**BC**), (2) those variables that are being held fixed (**S**), (3) those whose values are being varied through intervention (a set including but not always limited to X), and (4) those variables that are left free to respond to the changes under the given background conditions (a set including but usually not limited to Y).

According to this criterion, a hypothetical experiment relevant to the evaluation of the claim "X is a cause of Y" is one in which we hold some set of variables $\mathbf{S}$ fixed while intervening on X, and we observe any associated changes in the value of Y. The claim "X is a cause of Y" will be true if and only if changes would be observed in Y in the context of *some* hypothetical experiment defined by a specific $\mathbf{BC}$, $\mathbf{S}$, and I.

An important thing to note about this way of spelling out the meaning of a causal claim is that it makes use of a particular kind of counterfactual claim. In order to make sense of how an intervention on one variable, X, "makes a difference" to another variable, Y, we need to have some concept of what *would have happened* had the intervention on X not occurred. It is only by comparing the case in which the intervention is performed with our background understanding of what would have happened had the intervention not been performed that we get a sense of an effect.[7]

A second thing to notice about the hypothetical experiment referenced by a causal claim is that it involves two different types of interactions with the experimental system. The value of Y is observed, as may be the values of some of the background condition variables in $\mathbf{BC}$. Nothing is done to directly force these variables to take on particular values. For X and for the set $\mathbf{S}$, however, interventions directly force these variables to take on certain values. The distinction between *observing* the value of certain variables and *intervening* to set the value of others is absolutely central to the interventionist account of causation. The character of the knowledge that we gain from observing a natural course of events in a system and that of the knowledge that we can gain from carefully designed interventions on that same system are essentially different. Recall that Mill emphasized a related point (see section 2.4.1): one cannot know that an association between two variables is *not* a common-cause scenario until he or she has *produced* a change in the value of one by means of the other.[8] Section 3.1.2

[7] Woodward (2003b,a), Glymour (1986), Pearl (2009), von Wright ([1973] 1993), Holland (1986b), and Shadish et al. (2002) all recognize and comment on the counterfactual nature of causal claims within a interventionist framework.

[8] Menzies and Price (1993) take the experience of agency itself to be that which provides insight about a causal relationship between two variables, but such a view has since been heavily criticized, for good reasons. The knowledge unique to intervention need not be interpreted as stemming from any particular insight had by an experimenter as agent; instead, it comes from the experimenter's knowledge of the precise relationship that holds between his or her action and the system under study. For example, it would seem ridiculous to think that, if two collaborating scientists were to have identical, intimate knowledge of their experimental apparatus, only the one who performed the specific intervention involved in the experiment was the one to

below will discuss the particular characteristics that an intervention variable must have in relation to a system in order to allow us to distinguish between the situation in which X is a cause of Y and the situation in which it is not.

A few more items are in need of comment. How does the set $\mathbf{BC}$ figure in our characterization of the hypothetical experiment? The set of background conditions being considered is potentially relevant to the determination that a hypothetical experiment will provide on a causal question. For example, an experiment involving superconductivity must be carried out at an extremely low temperature—*i.e.*, one below the characteristic temperature of the conductor in question. Any causal relationships (or other relationships) being tested by the experiment will break down at higher temperatures, and may even be sensitive to changes in temperature within a certain range. We don't want to disqualify as causal any relationships that don't hold under *all* conditions, since such universal relationships will likely be rare (if they exist at all). Thus, the interventionist criterion requires only that there is at least one such set of conditions (more likely, there will be a range of conditions) under which the stated relationship (*i.e.*, that between an intervention on X and values of Y) holds for X to be considered a cause of Y.

Likewise for the set $\mathbf{S}$: for X to be considered a cause of Y, it is required only that there be *some* set $\mathbf{S}$ of variables that are held fixed while the intervention on X is carried out, as long as some difference would be seen in the value of Y under the conditions of the hypothetical experiment. But why is it even necessary to include the set $\mathbf{S}$ in our characterization of the hypothetical experiment? The main reason is that, particularly in the physical sciences, it is not enough to simply hope that all background conditions will remain constant across cases in which we intervene differently on X.[9] Instead, we must *ensure* that various potential influences on Y remain constant by *controlling* those particular variables or factors that we

take away the causal knowledge. The collaborator who did not perform the intervention obviously has an opportunity for causal learning equal to the one who did intervene; this knowledge is had by virtue of an understanding of the experimental setup and its relation to the intervention, not by the action of performing the intervention. Spirtes et al. (2000, Section 9.1) provides an interesting discussion of the situations in which "virtual" control variables, if they bear the same relationship to a causal system as an intervention and are known to have precisely that relationship, are epistemically equivalent to intervention variables themselves.

[9] Multiple-intervention experiments, aside from being necessary to ensure constancy of conditions, are also a tool for experimental efficiency (*i.e.*, reducing the total number of experiments required to learn a causal structure). See Eberhardt 2007.

suspect might influence the experiment. For example, in electrical experiments dealing with very low currents, electrical noise from power lines and AC wiring can disturb measurements, so a Faraday cage may be used to effectively *set* the ambient electrical field to zero. The set **S** can actually be quite large in a typical experiment in physical science; as we will see in chapter 5, interventionist control over multiple variables is quite important in thermodynamics.[10]

Another reason for consideration of the set **S** in the criterion above is more subtle. We want our criterion to correctly identify as causal cases in which X causes Y by two separate paths whose net effect is to cancel out.[11] Consider the classic example from Hesslow (1976) in which, conceivably, contraceptive pills could end up having no net effect on the likelihood of thrombosis, despite the fact that they actually influence the likelihood of thrombosis through two separate causal paths. In the absence of pregnancy, contraceptive pills increase the likelihood of thrombosis. However, contraceptive pills also decrease the likelihood of pregnancy, which in turn increases the likelihood of thrombosis. In this example, a hypothetical experiment in which an intervention is performed to stop a set of women from taking contraceptive pills could turn out to show no associated change in the incidence of thrombosis. If, however, we alter the design of our hypothetical experiment in such a way that we use some other means (perhaps a second contraceptive device) to effectively ensure that the incidence rate of pregnancy in our sample is zero (*i.e.*, we effectively hold pregnancy fixed), then we *will* see an association between our intervention on contraceptive pills and the incidence rate of thrombosis.[12]

Now why is there only the requirement that there be *some* intervention on X under which the change in value of Y is seen? As we mentioned above, there will usually be many different kinds of possible interventions on X. For example, if the causal relationship in

[10] In general, auxiliary interventions are more important in experiments in physical science, where we attempt to actively control the experimental circumstances. In social and medical sciences, however, the attempt to control the experimental circumstances is generally recognized as futile; auxiliary interventions become less important and the set **S** may be empty. When this is the case, randomization is generally used (see section 3.2.1 below).

[11] I am here referring to a particular case of "unfaithfulness", as it is called in the causal modeling literature. In cases of unfaithfulness, observational data (and even some experimental data) can make it appear that two variables are independent of one another despite one being a cause of the other. See Spirtes et al. 2000, 13–14; Woodward 2003b, 49–50; and Zhang and Spirtes 2008.

[12] Note that holding a variable in **S** fixed is actually an intervention, and thus is different from merely observing that the variable takes on a certain value. Experiments with non-empty **S** will be multiple-intervention experiments.

question is between soil quality and plant height, I might imagine a whole slew of hypothetical experiments with different kinds of interventions. One experiment might involve planting many of the same plants in my backyard and adding varying amounts of the fertilizer to each. Another experiment might involve planting many of the same seeds in completely different media (*e.g.*, some in manure, some in sawdust, some in the native soil of the area, some in coffee grounds, *etc.*). There are potentially many ways of performing an intervention on soil quality, and even when we consider the same type of intervention (*e.g.*, adding fertilizer), we can vary the magnitude of the intervention (*e.g.*, by adding more or less fertilizer). The criterion I have given does not require that a causal relationship be robust across *all* imaginable interventions on the purported cause variable, just *some*, although it is likely that there will be a range of interventions for which the causal relationship in question would hold, if it holds at all.

It is worth emphasizing, for the purpose of further clarification, that the criterion for causation that I have discussed here involves only an existential requirement. For X to be a cause of Y, there must exist *some* intervention I on X with respect to Y, *some* set of background conditions **BC** having *some* values, and *some* set of auxiliary variables **S** held fixed at *some* set of respective values such that a change is seen in Y. As I mentioned above, even when X is a cause of Y, there may very well be—in fact, there almost certainly *will* be—certain choices of I, **BC** and respective values $\{bc_1, bc_2, \ldots, bc_n\}$, **S** and respective values $\{s_1, s_2, \ldots, s_m\}$ such that no difference is seen in Y.

3.1.2 Ideal intervention

In order to ensure that the hypothetical experiment described in the criterion above reveals a *causal* relationship between X and Y and not a relationship of any other type, we must place careful constraints on what types of hypothetical actions might constitute interventions. The notion of *ideal intervention* is a conceptual tool that has been used by the interventionist account of causation for further limiting hypothetical experiments to just those that are legitimate and relevant for evaluating causal claims.

Several authors have given formal characterizations of an intervention or manipulation relative to a fully specified model of the causal system upon which the intervention is acting (*e.g.*, Spirtes et al. (2000), Pearl (2009), Eberhardt (2007)). Here, I will rely on the notion of ideal intervention given by Woodward (2003b), which does not rely on a fully specified model of the causal system. The advantage of Woodward's way of characterizing intervention is that it allows us to characterize what an intervention is from a perspective that is somewhat akin to that of the experimental researcher himself/herself—*i.e.*, a perspective which may include little or no prior knowledge about the about the causal structure of the system under investigation and which is oriented toward the goal of discovering some (or all) of that structure.

For Woodward, the notion of an intervention on a variable X only makes sense *relative to* another variable Y. Under Woodward's specification, an intervention is assumed to be undertaken for the purpose of learning about a causal system: specifically, for learning about one particular relationship within a causal system—*i.e.*, the relationship between X and Y. Because of this well-specified purpose, an intervention must satisfy several criteria in order to ensure that there are not confounding factors that might make it *appear* that X causes Y under the intervention, when such is not actually the case. The idea is that, under an intervention on X with respect to Y, an association should only remain between the two variables X and Y in the case that X is indeed a cause of Y. Thus, the constitutive criteria for an intervention on X with respect to Y must rule out all other possible sources of the association other than a causal relationship between X and Y.

Woodward defines an intervention in terms of an *intervention variable* taking on a certain value such that it causes X to take on its respective value. The following are Woodward's criteria for an intervention variable (Woodward 2003b, 98):

> What it is for I to be an *intervention variable* with respect to Y:
> I1. I causes X.
> I2. I acts as a switch for all other variables that cause X. That is, certain values of I are such that when I attains those values, X ceases to depend on the values of other variables that cause X and instead depends only on the value taken by I.
> I3. Any directed path from I to Y goes through X. That is, I does not directly cause Y and is not a cause of any of the causes of Y that are distinct from X except, of course, for those causes of Y, if any, that are built into the I-X-Y connection itself; that is, except for (a) any causes of Y that are effects of X (*i.e.*, variables that are causally between X and Y) and (b) any causes of Y that are between I and X and have no effect on Y independently of X.

I4. I is statistically independent of any variable Z that causes Y and that is on a directed path that does not go through X.

Each of the above criteria deserves comment. Criterion I1 is an acknowledgement that an intervention itself is a causal interaction, so that when the interventionist account characterizes what it is for there to be a causal relationship between X and Y in terms of an intervention on X with respect to Y, it is not attempting to do so in a non-circular manner. However, the characterization of the causal relationship between X and Y is spelled out in terms of a causal relationship that is, in a way, epistemically more familiar to us; we have far more experience concerning the results of our own actions and interventions than we do concerning other types of causal relationships.

Criterion I2 asserts that when certain values of the intervention variable obtain, it exerts complete control over the value of X. Although this criterion will not hold in many cases of actual intervention and experimentation (and as we will see below in section 3.2.2, need not), it gives clarity to the conceptual analysis of causation by removing any complications that would be involved if the intervention could not force a certain value on X. An intervention variable defined as having such complete control over the value of X allows for the conceptual space within which to entertain various counterfactual scenarios in which changes in the value of X are seen to correlate (or not correlate) with the value of Y. I2 is especially important in the case that X and Y have a common cause; if I exerts complete control over the value of X, then influence of the common cause on X is broken and the association induced between X and Y by the common cause will be eliminated, so that any remaining association can only be due to X being a cause of Y.[13]

Criterion I3 ensures that I has no influence on Y except by means of X. If I were to have a causal influence on Y through some other channel than through X, the association that we detect between our intervention on X and the value of Y might not be wholly due to a *causal* relationship between X and Y, but rather might be due simply to the correlation induced by I being a common cause of X and Y. For example, consider an experiment in which vitamin E injections are being tested for causal efficacy in treating children at risk for retrolental

[13] Inferences can still be made in cases where I2 does not hold. This possibility will be discussed briefly in section 3.2.2. See also Eberhardt and Scheines (2007) and Eberhardt (2007) on 'soft' or 'parametric' interventions.

fibroplasia. If the method by which the injections are administered also involves opening the infants' incubators and beneficially lowering barometric pressure and the infants' high blood oxygen levels, then the intervention has influenced the purported effect variable—*i.e.*, the health of the child—in a distinct way from the causal pathway being tested—*i.e.*, the influence of vitamin E on the health of the child.[14] In order to avoid this kind of problem, we stipulate that the intervention must not have this kind of secondary influence on Y that does not pass through X. This stipulation is what some causal theorists refer to as the "surgicality" of an intervention; a "surgical" intervention has causal influence on one and only one variable: the purported cause X.

We also must protect against a case in which our intervention induces an association between X and Y, not by causally influencing Y, but by being associated with a different cause of Y than X. Criterion I4 does this. Imagine, for example, an experiment that attempts to test a treatment that is thought to reduce the frequency and severity of the common cold. The researchers' advertisements for their study explain its goals and request participants. Suppose that, when deciding how to divide the subjects into test (*i.e.*, treatment) and control (*i.e.*, placebo) groups, the researchers choose to put the first 50% of the subjects to respond into the test group, and the later 50% of the subjects into the control group. Such a division into test and control groups could potentially induce an association between the intervention and the results of the study, if, for example, the earliest responders to the advertisements, who turn out to be the group receiving treatment, were people who already have a greater problem than average with severe and frequent colds.

For my purposes in this dissertation, I will simplify the above requirements down to two. I will allow that any experimental means of varying X can be considered an intervention on X, but for an intervention I to be adequate for the purposes of determining whether X is a cause of Y, it is important that it meet two requirements:

For I to be an *intervention on X with respect to Y*:

R1. Any factors that are potential causes of X (besides I) must be held constant across the variations we impose on X by means of I.[15]

[14] This example is taken from a discussion in Pearl (2009, 358).

[15] Woodward (2003b) requires (in his condition I2) that X cease to depend on its other causes when the intervention variable I takes on certain values. This requirement is intended to break any possible common-cause relationships between X and Y, but is too strict, especially in physics experiments where we usually

R2. I must not influence Y by any means other than through X.

Requirement R1 serves to ensure that it is I, and only I, which enforces the difference in the value of X. If a variation is detected in the value of Y, we want to be sure that it is entirely due to variation in X that was caused by our intervention. Any other source of variation in the value of X could potentially be a cause of variation detected in Y. Consider the classic barometer-storm example. If we wanted to test whether or not a barometer reading is a cause of a storm, we should not allow the air pressure to fluctuate, for fear that those fluctuations might bear a non-random relationship to our intervention in which we move the needle of the barometer.[16] Otherwise, it might appear that our intervention on the barometer's reading is a cause of a storm, when instead the barometer reading and the storm have a common cause (air pressure). Requirement R2 is important for distinguishing between a situation in which one variable causes another (*i.e.*, X is a cause of Y) and a situation where two variables share a common cause (*i.e.*, C is a cause of both X and Y). If our intervention influences Y by a means that is not through X, the intervention is a common cause of both variables, and the intervention provides no indication whatsoever of whether or not X is a cause of Y. For example, in a hypothetical experiment designed to test whether or not cholesterol is a cause of heart attack, the intervention in question must be one which influences the likelihood of heart attack *only* by means of cholesterol, rather than by some other means. Administering a drug which lowers cholesterol but also influences the likelihood of heart attack by lowering blood pressure would not be a suitable intervention. Distinguishing genuine causal relationships from mere associations due to common causes is one of the primary purposes of the interventionist account.

---

do not have the power to make a variable completely cease to depend on another. This weaker version of the requirement asks only that the other potential causes of X be held constant while X is varied by means of I. This type of intervention is sometimes called a "soft" or "parametric" intervention (see Eberhardt and Scheines (2007) and Eberhardt (2007)).

[16] Rather than holding all other potential causes of X constant, another methodological possibility is the use of a randomized intervention. In the limit of a large number of cases, the difference produced by variations in other potential causes of X is expected to average out. This solution is often used in the social sciences or medical studies, where the hope of controlling for all unknown causes is low.

3.2 FROM HYPOTHETICAL EXPERIMENT TO REAL EXPERIMENT

The conceptual tools and criteria discussed in the previous section serve the primary goal of the interventionist account of causation: that of explicating and interpreting causal claims in terms of hypothetical experiments. Given a causal claim, these tools allow us to reconstruct the relevant hypothetical experiment that would test the claim.

Although the conceptual interpretation of causal claims is the primary goal of the interventionist account, the interventionist account of causation carries with it an important corollary for scientific practice. For those who wish not only to evaluate the content of a causal claim but moreover to test its truth, the interventionist account can provide norms and recommendations for experimental testing. The truth or falsity of a causal claim can be empirically tested as long the hypothetical experiment embedded in the content of the claim can be actually realized. Actual experiments intended to test a causal claim can—and should—be modeled on the hypothetical experiment suggested in the content of the causal claim.

As a first step toward considering actual experiments for testing causal claims, let us adjust our language from the definition in section 3.1.1 above to focus on how an actual experiment must be carried out if it is to test a causal claim.

> Experimental instance for testing the claim "X is a cause of Y": Under some set of background conditions $\mathbf{BC} = \{BC_1, BC_2, \ldots, BC_n\}$ having values $\{bc_1, bc_2, \ldots, bc_n\}$, hold some set $\mathbf{S} = \{S_1, S_2, \ldots, S_m\}$ of variables other than X and Y fixed at values $\{s_1, s_2, \ldots, s_m\}$, perform an intervention I on X, and observe the value of Y.

What I have described here is only a single instance of an experiment and is insufficient for answering the question "Is X a cause of Y?" Recall that the hypothetical experiment embodied in the claim that X causes Y makes use of a contrast between two counterfactual states: the state of Y when X is manipulated in one way, and the state of Y if X *had been* manipulated in a different way (or not at all). But actual experiments provide us no access to such counterfactual knowledge; we need a way of estimating counterfactual results.

3.2.1 The "Fundamental Problem" of causal inference

The first problem we encounter in transforming our earlier criterion involving a *hypothetical* experiment into a methodological principle for testing causal claims using *actual* experimental instances is the following question: how we are to test the counterfactual claim involved in the criterion for being a interventionist cause? As noted before, evaluation of a causal claim necessarily involves a comparison between what happens under intervention and what *would* have happened in a different scenario (either under a different intervention or without the intervention). Thus, we are required to compare an experimental instance that we *did* test with an instance that we did *not* test.

Paul Holland (1986a) calls this problem the *Fundamental Problem of Causal Inference*. He states the problem in the language of his own framework, but I will paraphrase into the language I have been using here:

> FUNDAMENTAL PROBLEM OF CAUSAL INFERENCE: If there is a causal relationship from X to Y, it will only manifest itself in a relationship between a certain range of values of X enforced by intervention and a certain range of observed values of Y. Specifically, we are interested in the difference between the observed value of Y when we force certain values on X and what the value of Y *would* have been if we had forced a different value on X (or had we perhaps not intervened at all). However, for any one instance of the experimental system, we can only choose *one* intervention that forces *one* particular value on X, and we can only observe *one* value of Y for that instance.

As we discussed above, the claim being tested—*i.e.*, that X is a cause of Y—is a counterfactual claim. In order to test it, we need to make certain assumptions that allow us to transform our knowledge about *actual* events—*i.e.*, knowledge about the actual results of single instances of our experimental system—into *counterfactual* knowledge about our system—*i.e.*, knowledge about how our experimental system *would* behave, in general, if intervened upon in a certain way.

The obvious way to estimate the results of counterfactual experimental instances is to test many instances of the experimental system under similar conditions and to use statistical analysis to estimate the expected response of the system under different interventions.[17] Let

[17] I use "statistical analysis" here as an umbrella term for a wide range of methods that may vary in complexity. For my purposes, "statistical analysis" could be as simple as taking an average, or it could involve far more sophisticated techniques designed to identify a non-standard probability density function over the possible values of the response variable.

us define for this purpose an *experimental series*:

> EXPERIMENTAL SERIES FOR TESTING THE CLAIM "X IS A CAUSE OF Y": A set of two or more experimental instances for testing the claim "X is a cause of Y" such that:
> 1. Every instance in the set has the same (or sufficiently similar) values for **BC** and **S**; and
> 2. The set can be partitioned into two or more non-empty subsets such that every instance in each subset has the same value for the intervention[18] I on X and no two instances falling into different subsets have the same value for the intervention I on X.

As a simple example, consider the experimental series illustrated in figure 3.1. The series is a set of $j + k$ experimental instances partitioned into two subsets **A** and **B**. For each subset, a different kind of intervention on X is tested. For subset **A**, the intervention $I = i_a$ is used, which effectively sets X to x_a for each of the j instances in the subset. Likewise, for subset **B**, the intervention $I = i_b$ is used, which effectively sets X to x_b for each of the k instances in the subset.

The observations made of the value of Y for each of the subsets can be collated and used to generate a statistical estimate of the expected value of Y (or the probability distribution over the values of Y) for the type of intervention used in that subset of experimental instances. If there is a significant difference in the expected values of Y (or a significant difference in the probability distribution over the values of Y) for different subsets, then we may conclude that X is a cause of Y. If there is not a significant difference in the expected value of Y for different subsets, the conclusion must be more tentative. If a sufficient number of instances has been tested, we can legitimately conclude only that X is not a cause of Y under the particular circumstances of the experiment (where "circumstances" includes the background conditions **BC**, the choice of **S** on which to perform secondary interventions, and the range of values of X that were effectively tested in the series). The possibility that X will manifest itself as a cause of Y under other circumstances remains open, but the likelihood of that possibility can be reduced by testing of other series with different values for **BC**, different values for **S**, and/or interventions testing differing ranges of values of X.[19]

In order for a statistical analysis of the values of Y to legitimately represent (and be a

[18] The intervention in every experimental instance must meet requirements R1 and R2 discussed above.

[19] The schema I have given here is a more sophisticated version of that given by von Wright ([1973] 1993) for verifying counterfactual statements by experiment.

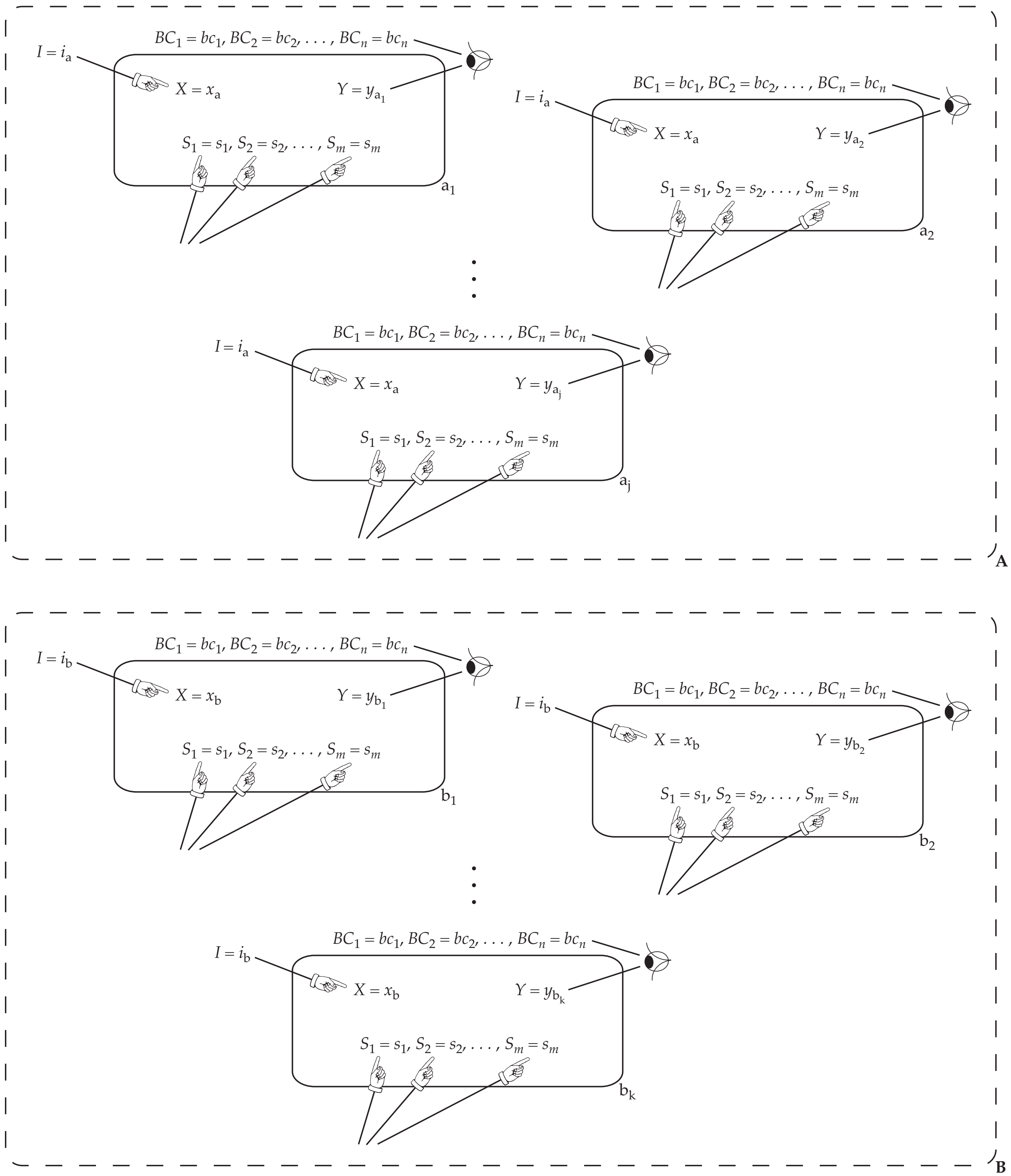


Figure 3.1: Illustration of multiple instances of an experiment and the interventions and observations involved in each. Pointing hands indicate intervention on the part of the experimenter, while eyes indicate observation without intervention.

predictor for) the expected value of Y under different interventions on X, at least one of two types of assumptions must be made.[20] One type of assumption is a *homogeneity assumption*: that the various instances of the experimental system are essentially identical, or at least similar "enough" to one another to behave in the same way under the same set of inputs. The following is a non-exhaustive list of homogeneity assumptions that are commonly made in the case of experiments in the physical sciences:

- *Time translation invariance.* The behavior of the system does not depend on the absolute time at which the experiment is carried out.[21]
- *Transience.* The behavior of the system is not affected by the previous occasions in which the experiment was run on the same setup. Each instance of the experimental system is temporally isolated from every other, or the effects of any other instance are deemed negligible.
- *Response similarity.* Each instance of the experimental system is similar in all relevant aspects; specifically, each instance will manifest the same Y for the same settings of all other variables in $\mathbf{BC}$, $\mathbf{S}$, and X, or will at least generate the same probability distribution over the values of Y for the same setting of the other variables.[22]
- *Isolation.* Among the set of experimental instances, there is no factor relevant to Y that varies (other than X).[23]

The other type of assumption, usually only applied when any of the above homogeneity assumptions do not appear to be valid, is a *randomization assumption*. The randomization assumption is most commonly used in the social sciences, when it is doubtful that the various

[20]Because my concern in this dissertation is on experimentation in physics, I concentrate on homogeneity assumptions here. For many social sciences, however, homogeneity assumptions will not be appropriate because the similarity among experimental units might be suspect. Holland (1986a) talks about there being two "types" of solutions to the Fundamental Problem, which he calls "scientific" and "statistical". For Holland, "scientific" solutions rely on assumptions about how the various instances of the experimental system are similar enough to each other to infer from the results of a certain type of intervention on one system to a likely result on another. "Statistical" solutions are used in cases where similarity between the various instances of the experimental system cannot be assumed; instead, randomization is used and *average* effects of different types of interventions are sought.

[21]This assumption can be weakened considerably; we might simply assume that the causal relations governing the system do not change over the time scale that is of interest to the experimenter.

[22]Holland (1986a) refers to this assumption about similarity of responses as "unit homogeneity", but I consider it as one assumption among many about the similarity of units.

[23]This assumption grounds the idea that the only thing that could have made a difference to the experimental outcome Y is the intentional variation of X by the experimenter.

instances of the system being studied are similar in the appropriate ways. The isolation assumption is a particularly strong assumption that is not likely to hold, for instance, of diverse national economies. Response similarity, also, is an assumption that is too strong for most psychological experiments, because different people in similar situations will not always have similar intentions or responses. Where homogeneity assumptions cannot be made, there is the risk that the application of different interventions will be systematically associated with differences in the various experimental units. Therefore, a different strategy must be used: instead of attempting to enforce unit homogeneity or assume it where it cannot be had, the experimenter determines the intervention that will be applied to each experimental instance by a random device. By randomizing the decision of what type of intervention to apply in each instance, the chances of an unintended association with other differences in the instances is lessened. And as the number of experimental instances grows, the differences that do exist among the various instances will begin to average out.

3.2.2 Non-ideal interventions

In section 3.1.2, I discussed the notion of ideal intervention, which is utilized in the description of hypothetical experiments that clarify the meaning of causal claims. Specifically, I discussed Woodward's four criteria (I1–I4) for an intervention variable. In the context of an actual experimental intervention, however, we will not always be able to achieve all four criteria. Moreover, we will not always know for sure whether or not we have achieved all four criteria.

Criterion I2, which requires that X cease to depend on any of its other causes, is too strict for many actual experimental interventions. For example, one may want to design an experiment to test if blood pressure is a cause of cardiovascular disease. There are several well-known interventions on blood pressure: for example, dietary changes, increased exercise, and medications. However, none of these interventions, whether implemented singly or together, will override or eliminate the influence of other causes of blood pressure. All of the other causes will remain in place, and we can only hope that the intervention will modify blood pressure from what it otherwise would have been without the intervention. There is no possible intervention that can arbitrarily "set" a person's blood pressure. An intervention that *modifies*—rather than sets—the value (or probability distribution over the values) of its target variable is called a "soft" or "parametric" intervention.[24]

The main disadvantage of an intervention that does not meet criterion I2, when compared to an intervention that does, is that the non-ideal intervention does not eliminate associations between X and Y that are due to a common cause. Under an ideal intervention on X, the mere presence of a continued association between X and Y is evidence for X being a cause of Y. This is not the case, however, for an intervention that does not meet criterion I2. The presence of an association between X and Y cannot alone be the basis for concluding that X is a cause of Y, because this association could simply be due to a common cause that remains intact under the intervention. Still, even when X does not depend *only* on the value taken by the intervention variable I (as specified in criterion I2), a properly designed experiment implementing the intervention can still allow for a conclusion about whether or not X is a cause of Y. As long as we can be confident that the intervention—which is exogenous to the system—is not associated with any other cause that is endogenous to the system (*i.e.*, as long as we are confident of criterion I4), then a significant difference in the value of Y seen for different values or kinds of intervention on X can allow us to legitimately conclude that X is a cause of Y. Where a significant difference is not seen, we can—analogously to the case of ideal intervention—make the tentative conclusion that X is not a cause of Y under the particular circumstances of the experiment (where "circumstances" must be taken to include the background conditions $\mathbf{BC}$, the choice of $\mathbf{S}$ on which to perform secondary interventions, and the range of values of X that were effectively tested). Again, the possibility that X is a cause of Y under other circumstances remains open, but the likelihood of that possibility can be reduced by expanded testing of background conditions and testing of a wider range of values for X.

[24] Eberhardt and Scheines (2007) and Eberhardt (2007) discuss soft interventions and their epistemic advantages and disadvantages in causal discovery when compared to "hard" or "structural" interventions that meet criterion I2.

3.3 FROM REAL EXPERIMENTS TO CAUSAL CLAIMS

Now that we have discussed the question of how to design an actual experiment to test a causal claim, I would like to consider the converse question: If we consider as given an arbitrary actual experiment, can we analyze it to say what kind of causal conclusions can be drawn from its results?

In asking this question, I mean to cast the net wide. I do not want to restrict attention just to those experiments that are explicitly designed for the purpose of testing causal claims or hypotheses. I wish to include anything that we might naturally call an "experiment"—*i.e.*, a scientific study in which the investigator deliberately sets up and/or intervenes on a system for the purpose of studying it. Purely observational studies (*e.g.*, observing astronomical events through a telescope, analyzing retrospective health information, *etc.*) that involve no intervention or set-up on the part of the investigator will not be considered experiments for my purposes here.

Despite the fact that, on my definition, all experiments will involve some type of intervention, we cannot expect it to be true that all experiments will provide a reliable answer to some causal question. Such an expectation would be trivially false, since it is obviously true that a bad assumption or a flawed experimental setup can lead to a failure to adequately answer a scientific question (whether causal or not). Nor can we expect that an experimenter will always have an explicit causal question in mind when designing or performing an experiment. Experiments in the physical sciences, in particular, rarely seem to be framed in terms of causal questions, at least not explicit ones. Consider an experiment aimed at measuring the boiling temperature of nitric acid at atmospheric pressure. Is such an experiment intended to test a causal claim? It certainly does not seem so, at least not at first glance. But could the experiment still afford causal inference, if we knew where to look and what assumptions to apply? I take the answer to this latter question to be non- obvious, and the goal of this section is to make some progress toward an answer.

This section attempts to apply the manipulationist account of causation to this broader range of experiments—a range that extends beyond the set of experiments that are explicitly designed for the testing of causal claims. In order to determine and make explicit any possible causal conclusions that can be drawn from a given experiment, I will first break down the structure of the experiment into three different types of components. I call *intervention components* those procedural steps in which the experimenter acts to ensure that a variable takes on a particular value or that a particular element of the experimental apparatus is set up according to some specification. An *observation component* is an instance in which no direct action is taken on the experimental system, but an observation is made about its behavior. Once intervention components and observation components have been identified, I will use the definition of an "experimental series" from section 3.2.1 above as a criterion for identifying the set of causal inferences afforded by an experiment. I will also argue that the conditions that suffice for causal inference obtain quite commonly, even among "ordinary" experiments that are not explicitly designed for the testing of causal claims.

3.3.1 Example: The Berti experiment

We have already commented a great deal on experiments that are intended to test causal claims, and the design considerations that will determine success or failure in their ability to afford causal inference. Here I will turn our attention to experiments that are *not* explicitly designed for the purpose of testing of causal claims. When analyzing an experiment that was not designed for the purpose of testing a causal claim, we simply seek to identify anything that could be properly described as an experimental series.

Consider as an example an experiment performed by Gasparo Berti, which aimed to decide a philosophical controversy surrounding the possibility of a vacuum and test Galileo's predictions about the maximum height to which water could be raised by suction. The experiment was most likely carried out sometime in the years 1642–1643 in the company of several active participants in the scientific scene of Rome, including Raffaello Magiotti, Athanasius Kircher, and Niccolò Zucchi.[25] A description of the experiment is found in a 1648 letter from Magiotti to Marin Mersenne. The following is an excerpt from the letter:

[25]Drake (1970) gives this date estimate on the basis of textual clues in Magiotti's letter and Thomas Cornelius' *Progymnasmata physica*, the dates of communication between Magiotti and Torricelli, and the dates of Torricelli's visit to Rome. Drake differs from de Waard, who estimates that the experiment took place between 1639 and 1641 (see de Waard (1936, 180)). See (Middleton 1963) for information on the participants and their written accounts of the experiment.

In regard to the history of quicksilver, you may know that the many wells of Florence, which are cleaned each year by suction with siphons, gave Sig. Galileo the opportunity to observe the height of the attraction which was always the same, about 18 Tuscan *braccia*,[26] and that in every siphon or cylinder, no matter how wide or thin. This was the origin of his speculations on the subject in his work on the cohesion of solids.

Later, Sig. Gasparo Berti, here in Rome, made a lead siphon that stretched about 22 *braccia* from his courtyard to his room, and was filled from above in the following way. First, leaving both valves open (D below and F above), vessel AG was filled with water.[27] Then, after closing valve D, the water of vessel AGPM was poured out (through valve M), leaving the water inside the siphon at height AE. Later, making sure to keep vessel HE full,[28] the water AE was allowed to flow out through valve D, which (since valve F was already open and immersed in water) pulled the water from above and filled the whole siphon BA and the vessel AG. Finally, with vessel HF full and having closed valve F, and with vessel AG full (having first closed M) and D open, the water started to descend through the siphon, emptying the entire neck BF. The water continued to fall until reaching N and did not descend further, but almost always balanced itself [at N] when the experience was replicated. And it was possible to observe this very well, since part BC of the siphon was made of glass on purpose and the whole siphon was well glued and watertight. Sig. Berti believed that he could refute Sig. Galileo with this experience, saying that the length from N to A was more than 18 *braccia*, but he should have seen that the piece of the siphon AE doesn't count, being immersed in the water of vessel AG; EN was 18 *braccia* exactly.

I should not fail to mention one thing that gave me much to think about: while the water of the siphon was falling and the neck BF was emptying, an infinite number of tiny bubbles, like those in glasses and crystals, could be seen rising through the water inside the glass BC: this, without a doubt, was some stuff that went to refill where the air was missing. I could not convince myself that it was air because there was not enough air in the water in vessel AG to refill that space (besides, the space NBF could be made much larger and it would still refill). Nor could air have entered through pores or the welding of the siphon, for if it had, it would have eventually allowed the suspended water to fall. In fact, those bubbles have always remained in my mind: I can only explain my whole sentiment about them briefly like that. Now I wrote about this experience to Sig. Torricelli, believing that if the water had been seawater and thus heavier, it would not have stopped at N, but lower. They did the experience and arrived finally at quicksilver, Sig. Galileo's predictions always remaining consistent with the new speculations and experiences of Sig. Torricelli.[29]

Besides Magiotti's letter, there are four other authors who describe Berti's experiment: Zucchi (1648), Kircher (1650), Maignan (1653), and Schott ([1664] 1687).[30] These other

[26]The conversion for a Tuscan (Florentine) *braccio* prior to 1782 is given by Martini (1883) as either 0.551 meters for the *braccio a terra* or 0.584 meters for the *braccio a panno*.

[27]See figure 3.2.

[28]This was presumably done by continuous refilling. See figure 3.3, which depicts a different version of Berti's experiment in which the upper vessel was perched on a windowsill where it could be filled from a room on an upper floor.

[29]This English translation is mine. The manuscript of the letter, which was written in Italian, is published in de Waard (1936, 178–181).

[30]Zucchi (1648) was first published anonymously but later republished under his own name in a larger work. de Waard (1936) contains relevant excerpts (in the original Latin) from all four of these works. An

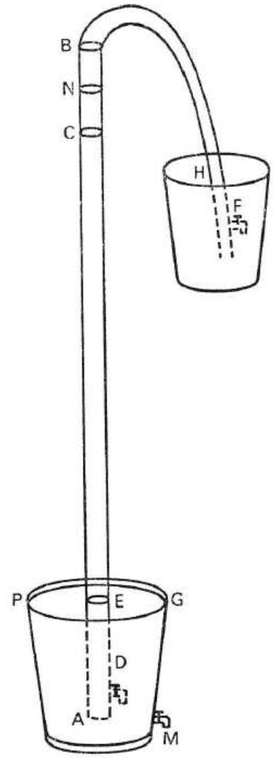


Figure 3.2: Diagram of Berti's experiment, included in Magiotti's letter to Mersenne. From Tannery et al. (1986, 169). Reproduced here by permission of Éditions du Centre National de la Recherche Scientifique.

accounts all describe a similar and slightly more complex version of the experiment, which may have been a later modification.[31] In this version, a glass globe was mounted on the siphon (see figure 3.3). The globe contained a bell attached to a magnetic device so that, once the purported vacuum was achieved, the bell could be rung from outside by using another magnet.

The primary intention of the experiment, at least on Berti's part, appears to have been a desire to check (and perhaps refute) Galileo's prediction of 18 *braccia*. A secondary intention was to investigate the empty space itself: was it or was it not a vacuum? It is obvious from Magiotti's letter that this latter was a question of interest for him, and it was likely the most

English translation of relevant excerpts of Maignan (1653) is given in Middleton (1963).

[31]Middleton (1963) assumes that the different accounts all referred to the same experiment, and assumes that Magiotti's memory was failing him when he wrote of the simple version in his letter. Drake (1970), on the other hand, takes the the differing descriptions to be of two or more distinct experiments, and I concur.

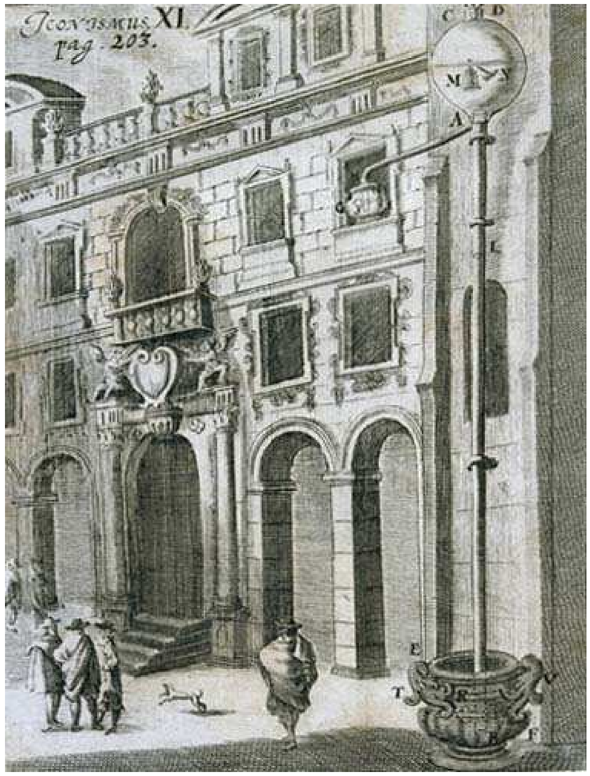


Figure 3.3: Engraving of a more complex version of Berti's experiment, reproduced in Schott ([1664] 1687, 203)

important question in the minds of the other participants as well; Zucchi and Kircher were both Jesuits who were convinced of the impossibility of the vacuum.

The addition of the bell in the more complex version of the experiment was suggested by Kircher and intended as an *experimentum crucis* to test the claim that the space in the globe was a vacuum. The space was found to transmit both light and magnetism, and the bell could indeed be heard when rung. These facts were enough to convince both Zucchi and Kircher, and perhaps also Berti, that the space was not a vacuum. Maignan, on the other hand, proposed that the sound of the bell was being conducted by the bell's wooden support rather than by the space itself, and was of the opinion that the space was indeed a vacuum. It seems that Magiotti remained uncertain. Inasmuch as the various participants walked away from the experiment with different views, the *experimentum crucis* was a failure.

Notice that the questions of interest for those performing and attending the Berti experiment were not causal questions; none of the writings explicitly mention a curiosity about the cause of the empty space, for example, nor is there any evidence of debate among the participants about what caused the elevation of the water to be 18 *braccia* rather than some other height. The questions posed and debated were, instead, factual questions and questions of interpretation about the phenomenon: How high did the water stand? Could there be any pores or imperfections in the device? Did the space transmit sound? Was the space a vacuum, or was it not?

Despite the lack of interest in causal questions on the part of those involved in the experiment, can causal conclusions can be drawn anyway? A first step toward deciding this question is to itemize the procedure described in the excerpt from Magiotti's letter and classify each item as an intervention component (I) or an observation component (O):

1. (I) Construct and set up the pipe and vessels in the configuration given in figure 3.2. Ensure that valve M is closed.
2. (I) Open valves D and F.
3. (I) Fill vessel AG with water.
4. (I) Open valve M.
5. (O) Observe that vessel AG empties. Water inside the siphon remains at height AE.
6. (I) Fill vessel HF with water.
7. (I) Open valve D and continue supplying HF with water.
8. (O) Observe that the water flows out through valve D and also flows from above to fill siphon.
9. (I) Close valve F and valve M.
10. (O) Observe that the water begins to descend down the siphon, emptying neck BF and falling until it reaches N.

And assuming a similar set-up for the more complex version of the experiment,[32] we might simply modify the first step and add several steps to the end of the procedure:

[32] The other accounts of the experiment describe a different procedure for filling the apparatus with water, but the difference in procedure is inconsequential for the analysis I offer below.

1*. (I) Construct pipe mounted with glass globe and internal magnet-bell apparatus. Arrange it and vessels in the configuration given in figure 3.3.

⋮

11. (O) Observe that light passes through the sphere.
12. (I) Move magnet around the exterior of the glass globe.
13. (O) Observe that the interior magnet moves in response to the exterior magnet's movement.
14. (O) Observe that sound can be heard from the bell inside the glass sphere.

It is interesting to notice that many—not just one—of the items listed above are interventions on the experimental system. Most of them serve only as steps toward the set-up of the apparatus. However, each can, in principle, be considered as an intervention in an experimental instance for testing a variety of causal claims; the variable X will be the thing intervened upon (for example, the intervention in step 4 is an intervention on whether or not valve M is open), the variable Y can be any observation that follows (for example, the observation in step 5), and all other observations and interventions involved in the experiment are considered either as observed background conditions in $\mathbf{BC}$ or auxiliary interventions in $\mathbf{S}$.

The question of whether or not the experiment affords causal inference will amount to the question of whether or not the various experimental instances that make up the experiment are part of an identifiable experimental *series*. Consider, for example, an experimental instance centered around the intervention in step 4 above. The variable X might represent the state of valve M (open or closed) and the variable Y might represent the state of the vessel AG (which can be empty or full, but is observed as empty in step 5). The set-up established in steps 1–3 and other background conditions surrounding the experiment could all be represented by the set $\mathbf{BC}$. Now, if we can identify at least one other experimental instance with the exact same values for $\mathbf{BC}$ but a different intervention on valve M, we will have identified an experimental series for testing the claim that the state of valve M is a cause of the vessel AG emptying. Berti's experiment does in fact provide such an experimental instance. Assuming that there is some time lapse between the execution of steps 3 and 4, we can consider as a second experimental instance the time period after steps 1–3 have been performed but before valve M has been opened. In this time period, vessel AG is observed to be full. Since there is a difference in the state of vessel AG between the former experimental instance in which M is opened and the latter experimental instance in which M is not opened, we can conclude that the state of valve M is a cause of the state of vessel AG.

The observation-intervention pair considered in the example experimental series just given (*i.e.*, a valve being opened and a vessel emptying) are such an ordinary matter of course that we do not tend to think of it as the basis for a causal conclusion that can be drawn from the experiment. That water only empties from a vessel that has some open outlet is a mundane fact that each person experiences so many times in life that it becomes an implicit piece of causal background knowledge. Still, inasmuch as an experiment establishes a contrast between performing and not performing an intervention (or alternatively, performing one type of intervention versus performing a different type of intervention) and the corresponding difference in the observations made in each case, the experiment will afford the conclusion that the one variable (the variable intervened upon) causes another (the variable observed to covary with the variable intervened upon).

But are there more substantial causal questions that could have been answered by the experiment in question? The interventions performed in the more complex version of the experiment, if compared to a relevant contrast case, could be interpreted as tests of causal questions. For example, when it is observed in step 11 that light passes through the spherical glass vessel, the implicit contrast case is whether or not light passes through the spherical glass vessel when it is filled with ordinary air. Presumably there were no noticeable differences between the appearance of images viewed through the vessel in the two cases. Likewise, we might compare the observations in steps 13 and 14 when they are made in the context of the experimental set-up and when they are made in a contrasting context (for example, with a column of water filling the siphon up to mark N, but not brought about through suction, so that the spherical glass vessel is filled with ordinary air). One might ask, for example, if there is any difference in the volume of the bell in each case, and then attribute any difference in volume to the difference between ordinary air and the purported vacuum inside the bell.

The participants in the experiment were not, however, thinking in terms of these contrasts. Even in they had been, they would have been unable to agree on a causal conclusion because they were unable to agree about what the interventions in the experiment had achieved. Recall that the causal question answered by an experiment is whether or not X causes Y, where X is the variable intervened on and Y is the variable observed. It clear in Berti's experiment what the intervention is (or rather, what the sequence of interventions is: items 1–4, 6–7, 9, 12) but what those interventions achieve was a subject of debate. Some of the participants—the vacuists—thought that those interventions achieved a vacuum in the spherical vessel, while others—the plenists—thought that the vessel was still filled with some sort of attenuated matter. If they had been able to agree, for example, that there was a vacuum in the vessel, then they might have been able to agree that the difference between vacuum and ordinary air was not a cause of the transmission of light or magnetism. In addition, they would have been able to reach a conclusion about the effect of the vacuum on the transmission of sound by noting the difference in the volume of the bell's ring in each case.

But there was no such an agreement. Instead, some of the participants were *already* certain, prior to the experiment, that a vacuum could not transmit light or sound or magnetic phenomena. They took themselves to be certain of the causal relationships, and they attempted to test the presence or absence of the vacuum by the presence or absence of these effects. An experiment which could have been understood as testing various causal claims instead used a prior confidence in those causal claims to test whether or not the cause factor was present.[33] Even so, the actual use to which the experiment was put does not prevent anyone who is later informed of the details of the experiment from drawing causal conclusions.

3.3.2 When does an experiment afford causal inference?

We have seen in the Berti experiment discussed in the previous section that an experiment can afford causal inference even if it was not originally designed or intended to test a causal claim. But how can we generalize this result to identify which experiments afford causal

[33] To be clear, I do not object to the use of prior causal knowledge in the interpretation of experimental results; reliance on prior causal knowledge will often (if not always) be necessary. However, in this case, the presumed causal knowledge had not been established through experiment via the methodology I have outlined in this chapter.

inference and which do not?

The criterion is extremely simple. Those experiments that afford causal inference will be exactly those experiments for which we can identify a set of instances that satisfies the definition of an *experimental series for testing the claim that "X is a cause of Y"*. We must, in other words, identify a set of instances which:

1. have identical set-up and background conditions;
2. have identical secondary interventions (possibly none at all);
3. all implement a single intervention upon the same variable X, but do so in at least two different ways; and
4. all observe some other variable Y.

Even among experiments that were not designed for the purpose of causal inference, many will still afford causal inferences. The requirements I have placed on an experimental series for testing a causal claim will be found quite commonly in "ordinary" scientific experiments. We can see that this is true especially when we consider that, if there is a time lapse between the set-up of the experiment and the intervention on the purported cause variable, and if the time latency of the observed result is small in comparison to the time lapse, then a comparison of observations made before and after the intervention is performed will usually correspond to an experimental series for testing if the variable intervened on is a cause of the subsequent observation.

In turning to the question that is the major focus of this dissertation, I now ask: what does this mean for experimental physics? At this point, we can at least assert that there will be a great *potential* for causal inference in most physics experiments. The extent to which this potential for causal inference is *used* is a question that I will now begin to treat by examining the history of thermodynamics as a case study. I will turn particular attention to the question of what *significant* (*i.e.*, not common sense) causal inferences were drawn from experiments and actually utilized by scientists in the history of thermodynamics, and whether or not these causal inferences informed subsequent theory.

4.0 THE EARLY HISTORY OF THERMODYNAMICS: A CASE STUDY IN CAUSAL REASONING AND THE RELATIONSHIP BETWEEN THEORY AND EXPERIMENT

> *Scientific activity, as we know it, consists of two basic components: observations and interventions. The combination of the two is what we call a* laboratory*, a place where we control some of the conditions and observe others. It so happened that standard algebras have served the observational component very well but thus far have not benefitted the interventional component.*
>
> Pearl (2009, 421)

In this chapter, I examine the early history of experiments in thermodynamics. In telling this history, it is easy to focus on those discoveries that are commonly recognized as the main theoretical results that came out of this period of experimentation: namely, a newly discovered property of the air (its "spring" or pressure) and Boyle's law (a quantitative formulation of the relation between the volume of the air and this newly-discovered property). In my account of the history, however, I will highlight the debates about causation and the use of actual experiments and hypothetical experiments to settle these debates. I will show that the discoveries of the period were made against a rich background of causal debate and arguments using causal-experimental contrasts. Specifically, I will show that the experiments, and the arguments of the scientists themselves about their own experiments, made use of experimental series matching precisely the schema for testing causal claims that I characterized in chapter 3.

In sections 4.1–4.3, I will describe the experiments of Torricelli, Roberval, and Pascal, highlighting the causal questions on which their experiments were brought to bear. In section 4.4 I discuss the experiments and theorizing of Boyle, who was the first of this series of experimentalists to attempt a quantitative measure of the phenomena surrounding the Torricellian apparatus. In the context of Boyle's work, I explore the relationship between experiment, mathematical representation, and physical theory. I begin by discussing what occurs when we turn our attention from the consideration of concrete experiments to more abstract theoretical considerations. Using Boyle's experiments with volume and pressure as my main example, I will argue for the thesis that certain pieces of knowledge that are salient in the experimental process become hidden or latent at the theoretical level as we abstract away from the details of a given experimental context. I will also argue that causal knowledge is one type of knowledge (among others) that becomes hidden in the process of abstraction. We cannot "read" causal relationships off of the mathematical formulas that are the result of the process of abstraction, let alone determine what account of causation our mathematical formulas might support. I will show, however, that by closely examining the generation of the mathematical formulas that we utilize in physical theory—*i.e.*, the experiments that constitute evidence for them and the abstraction process by which we arrive at them—we can identify their decidedly interventionist underpinnings. We can also explain how interventionist claims made in the context of a physical theory can be epistemically justified by reference to the experimental interventions and observations that serve as evidence for the theory.

4.1 TORRICELLI'S EXPERIMENTS

The use of what would later become known as the "Torricellian apparatus" began as a direct result of Galileo's speculations about the "force of the void" in the *Two New Sciences* (Galilei ([1638] 1989)).[1] As I discussed in section 3.3.1, Gasparo Berti, a mathematician and natural philosopher in Rome, was the first to successfully carry out an experiment to test Galileo's predictions about the maximum height of water that could be suspended by a vacuum. When Evangelista Torricelli heard about the Berti experiment, he began

[1] Refer back to section 2.2.1.3 for a discussion.

to hypothesize about more general principles that might govern different liquids and the behavior of the apparatus. In this and the next two sections, I will describe the historical sequence of experiments performed by Torricelli and others on his apparatus.

Magiotti mentioned in his 1648 letter (see excerpt in section 3.3.1) that he had written about the Berti experiment to Torricelli.[2] According to Dati (1663, 20), who was a student of Torricelli, Torricelli had been considering Galileo's arguments in the *Discorsi* about the maximum height of water that could be raised by suction, and had the idea of using mercury in place of water; because of the greater density of mercury, it would be a much more convenient experiment with a smaller apparatus.[3] Dati relates that Torricelli shared his idea with his friend Vincenzo Viviani, who was the first to carry it out.[4]

Regardless of how the idea of using mercury occurred to Torricelli, there is a striking contrast between Torricelli's purposes in performing his experiments and the purposes of Berti and his collaborators. In contrast to Berti and his collaborators, Torricelli's primary interest was in answering a causal question: what causes the water (or other liquid) to stand above the level in the vessel, and why at a certain height rather than another? Galileo had given his own account of the cause of the maximum height to which water can be elevated by suction; he did so by considering the column of water as a special case of cohesion in which only the force of a vacuum held the matter together. Torricelli had a completely different theory of the cause, which he relates in a 1644 letter to Michelangelo Ricci, of which the following is an excerpt:

> I already mentioned to you that a kind of philosophical experiment was being done regarding the vacuum, not simply in order to create a vacuum, but to make an instrument that might display the changes in the air, which is sometimes more heavy and thick, and other times more light and thin. Many have said that a vacuum cannot be made, and others that it can, but with the aversion of Nature, and with effort; I don't know that anyone has yet said that it does so without effort, and without any resistance from Nature. I reasoned as

[2]Of the surviving letters between Magiotti and Torricelli, none makes any mention of the Berti experiment. The surviving letters between the two men span the period from 1640–1644 and can be found in Loria and Vassura (1919). The subject of the letters are mostly geometry and Torricelli's art of lens grinding.

[3]It is interesting that the comment in Magiotti's letter about Torricelli settling on mercury follows immediately after his mention of the idea of trying seawater. The implication may be that Torricelli tried progressively denser liquids until arriving at mercury.

[4]It is also possible that it was Viviani's own idea to try the experiment using mercury. In a manuscript copy of Galileo's *Discorsi*, next to the discussion of the column of water breaking, Viviani wrote a comment about his expectation that other liquids would also rupture, but at differing heights in reverse proportion to their specific gravity. See Middleton (1963, 17–19).

follows: if one were to find a very manifest cause of the resistance that is felt in making a vacuum, it would be vain, it seems to me, for someone to seek to attribute to the vacuum that operation that openly derives from another cause. On the contrary, making certain very easy calculations, I find that the cause which I allege (*i.e.*, the weight of the air) should on its own have a greater effect than it does when attempting to make a vacuum. I say this so that some Philosopher, seeing that he cannot escape the admission that the weight of the air causes the resistance that one feels in the forming of the vacuum, might not say that he concedes the operation of the weight of the air but persist in affirming that Nature also contributes an aversion to the vacuum.

We live submersed at the bottom of an ocean of elemental air, of the kind which by indubitable experience is known to have weight, enough that the thickest air close to the surface of the earth weighs about one four-hundredth the weight of water. Authors have observed that at twilight, the vaporous and visible air rises above us about fifty, or maybe fifty-four miles; but I don't believe this, because it would show that the vacuum should produce a much greater resistance than it does, although there is the reply that the weight given by Galileo is intended to refer to the lowest air where men and animals live, but that above the peaks of the high mountains, the air begins to be very fine and has a much smaller weight than one four-hundredth the weight of water.

We have made many glass vessels like the following labeled A and B [see figure 4.1], with a neck two *braccia* in length. When we filled each with quicksilver and then, with a finger sealed over the mouth, turned them over in another basin containing quicksilver C, they could be seen emptying. The neck AD, however, always remained full to a height of one *braccio* and a quarter, and one *dito* more. In order to show that the [top of the] vessel was perfectly empty, the basin placed underneath was filled with water up to D, and raising the vessel little by little, when the mouth of the vessel arrived at the water, the quicksilver could be seen descending from the neck and refilling itself with water completely up to E, with a horrible impetus.

While the vessel AE was empty, and the quicksilver was held up in the neck AC even though extremely heavy, we discussed this force that holds the quicksilver against its natural tendency to fall back down. It has been believed until now to be internal to the vessel AE, either by vacuum, or by that extremely rarified substance;[5] but I claim that it is external, and that the force comes from outside. On the surface of the liquid in the basin weighs the height of fifty miles of air; so what marvel is it if, inside the glass CE, where the quicksilver has no inclination nor even repulsion (since nothing is there), it might enter, and raise itself until it balances itself with the weight of the external air, which pushes it?

Water, in a similar but much longer vessel, will rise almost to 18 *braccia*—as many times higher than quicksilver as quicksilver is denser than water—in order to balance itself with the same cause that presses on both the one and the other. The experience made at the same time with vessel A, and with the pipe B in which the quicksilver always rested at the same marked line AB, almost certainly confirmed the explanation that the force is not inside; since vessel AE, where there was more rarefied substance, should have had a more

[5]Torricelli's comment here about rarified substance (literally, rarified "stuff" [*quella roba sommamente rarefatta*]), both here and in the paragraph below, is likely satirical. It is a nod to philosophical arguments about whether or not the apparently empty space is a vacuum or instead filled with some rarified substance. Torricelli is himself a vacuist, as he shows by his experimental demonstration of water filling the empty space, but the difference between the vacuist and plenist positions matters little to him because of the positive causal theory he advances here.

vigorous and attractive force because of its greater rarefaction than that of the very small space B.

I then tried, with this principle, to account for all of the sorts of aversions attributed to the vacuum that are found in various effects, and until now I have not encountered anything for which it cannot account. I know that many objections will come to your mind, but I also hope that you will overcome them by thinking about them.

My principal intention—*i.e.*, to find out, with instrument EC, when the air is thicker and heavier and when it is thinner and lighter—was not able to succeed, since the level AB changes from other causes which I never would have believed—*i.e.*, by cold and heat, and very noticeably [*sensibilmente*], just as if the vessel AE were full of air.[6]

Figure 4.1: Diagram of Torricelli's experiment, included in his letter to Ricci, 11 June 1644. From Loria and Vassura (1919, 187).

In this letter, Torricelli proposes a new theory of the cause of the liquid rising in a vessel: it is not any force of the vacuum that raises the liquid, but rather the positive force of the weight of the air pressing down on the surface of the liquid exterior to the vessel. The weight of the raised liquid balances out the weight of the air, as in a balance.

Yet Torricelli does not give a positive argument for his causal account; presumably, he does not yet have the right kind of experimental evidence to prove that the weight of the air is the cause of the liquid rising. His main argument here is a negative one: that the force of

[6]My translation. Original text of the letter can be found in Loria and Vassura (1919, 186–188).

the vacuum is *not* the correct cause. He makes this argument by describing an experimental series for testing the claim that the force of the vacuum is the cause of the liquid's rising.

Recall the definitions for *experimental instance* and *experimental series* that I gave in the previous chapter (see section 3.2). Torricelli's argument makes use of an experimental series. He describes two experimental instances:

1A. Vessel A is made to have a neck two *braccia* in length and a large bulbous end. Vessel A is filled with quicksilver, stopped with a finger, and upended in a basin of mercury. The quicksilver is observed to fall to a height of one and a quarter *braccia* plus one *dito*.

1B. Vessel B is made to have a neck two *braccia* in length but with a smaller and more confined end than vessel A. Vessel B is filled with quicksilver, stopped with a finger, and upended in a basin of mercury. The quicksilver is observed to fall to a height of one and a quarter *braccia* plus one *dito*.

His experimental argument concerns the claim that the force of the vacuum is a cause of the mercury's rising. He explains that his two experimental instances in vessels A and B were performed "at the same time", thus assuring Ricci that there were similar—if not identical—background conditions **BC** (*e.g.*, time of day, temperature, altitude, "thickness" of the air, "heaviness" of the air). He also describes the identical procedure of secondary interventions **S** followed in both instances: each vessel was filled with quicksilver, sealed with a finger, and then upended in a larger container of quicksilver. His two experimental instances form an experimental series because they have similar background conditions **BC** and an identical procedure of secondary interventions **S**, but they also differ in one salient respect: the two vessels were purposely constructed differently so that vessel A had a larger bulbous space than did vessel B. Thus, the purposely-constructed difference in the vessels is a difference in the intervention I on the size and shape of the vacuum-space that results in each vessel.

Torricelli uses this experimental series of two instances as evidence *against* the claim that the force of the vacuum is a cause of the mercury's rising. What he describes is essentially a causal experiment with a null result. Despite intervening to ensure that there is a difference in the size and shape of the empty space above the level of the quicksilver in the two experimental instances, the quicksilver still rises to exactly the same height in both cases. If the force of the vacuum caused the quicksilver to rise, Torricelli argues, a difference in the size and shape of the space occupied by the vacuum (and any possible difference in its attractive force) would be detected as a difference in the height of the quicksilver. Since no such difference in the height of the quicksilver is detected in the two cases, Torricelli concludes that the force of the vacuum is *not* a cause of the mercury rising.[7]

The reason that Torricelli does not make a positive argument for his causal account is that he has not yet conceived of an experiment that would allow him to vary the positive weight that the air exerts on the liquid. In order to prove that the weight of the air is the cause, he would need to vary its weight and show that the height of the column varies.

Ricci responded with three objections to Torricelli's causal theory, of which I will discuss two:

> ...First of all, it seems to me that one could exclude the action of the air in gravitating on the outer surface of the quicksilver in the basin by placing on this vessel a cover pierced with a single hole through which the glass tube passes and then stopping all parts completely so that there would be no further communication. The air above the basin would, in such a case, gravitate no longer on the surface of the quicksilver, but upon the cover; and if then the quicksilver remains suspended in the air as before, the effect could no longer be attributed to the weight of the air which is supposed to hold it there in a sort of equilibrium. Secondly, if we take a syringe (which should be frequently used in this sort of inquiry and should have its sucker completely enclosed so as to exclude with its bulk every body), and if then we stop the hole on top and pull the piston back by force, we feel a great resistance; and that effect follows not only when the syringe is held downwards so that the sucker is brought above and the air gravitates upon the top of the piston rod, but it follows in whatever direction the syringe is turned. In these cases, it is still not evident that one can easily imagine how the weight of the air has anything to do with the effect (Cioffari 1937, 166–167).

Notice that Ricci uses the exact same form of experimental-causal argument as did Torricelli. Both of the experimental series that he suggests provide a contrast between an instance in which the weight of the air is pressing (vertically) down on the surface of the liquid and one in which the weight of the air is not pressing (or at least not pressing down

[7]As discussed in sections 3.1.1 and 3.2.1, null results in experimental series are *evidence for* the claim that X is not a cause of Y, but they cannot conclusively prove the claim. Torricelli appears to have been more certain of his negative claim than he perhaps should have been.

vertically), yet no difference is seen in the effect. In the first series that Ricci suggests, he contrasts Torricelli's experimental setup with one in which we instead place a rigid cover on the basin of mercury, such that the rigid cover supports the weight of the atmosphere, rather than the surface of the liquid itself. So this experimental series is made up of the following two instances:

2A. The Torricellian apparatus in a basin of mercury exposed to the atmosphere.
2B. The Torricellian apparatus in a basin of mercury, then sealed with a lid.

Ricci's argument can be put into the following form. These two instances constitute an experimental series because they could be performed under the same set of background conditions, they would involve the same procedure of auxiliary interventions, and they would differ only with respect to the intervention of sealing a lid onto the basin. If, then, the height of the mercury is the same in both instances, it would be evidence that the weight of the air is not a cause of the rising of the mercury.

The second contrast he sets up is one in which the plunger of a sealed syringe is pulled back with the syringe in two differing orientations:

3A. The hole of a syringe is stopped. The syringe is then held with the plunger in a horizontal orientation and pulled back.
3B. The hole of a syringe is stopped. The syringe is then held in a vertical orientation with the plunger pointing up. The plunger is pulled back.

Again, these two instances constitute an experimental series because they could be performed under the same background conditions, they would have the same set of auxiliary interventions, and they would differ only with respect to the intervention that orients the syringe and plunger. This difference is significant because Ricci assumes that the weight of the air can only push *down*, so any force on the piston due to the weight of the air must differ in the two orientations. Since no difference in the resistance of the piston is felt between the two cases (as Ricci assumes from common experience), the two instances are evidence that the weight of the air is not a cause of the resistance we feel when pulling the piston back.

Ricci's two experimental-causal arguments are well-justified objections to Torricelli's theory as he expressed it in the first letter. They directly concern Torricelli's positive claim that the weight of the air is the cause of the mercury rising. In Torricelli's response to Ricci's objections, he modifies his causal claim somewhat.

> I will respond to the first: If you bring in a soldered metal sheet that covers the surface of the basin, placing it in a manner so that it touches the quicksilver in the basin, then the raised quicksilver in the neck of the vessel will remain as before, not because of the weight of the sphere of the air, but because the basin cannot give it space. If instead you place the sheet so that it also captures some air inside, I ask if, of that enclosed air, you mean it to be of the same degree of compression as that outside, in which case the quicksilver will remain as before (by the example that I will give presently of wool), but if the air that you include is more rarefied than the external air, then the raised metal will descend a certain amount; if it were infinitely rarified—*i.e.*, vacuum—then the metal would descend all the way, as far as the enclosed space can allow it.
>
> The vessel $ABCD$ [see figure 4.2] is a cylinder full of wool or some other compressible material (let's say air). The vessel has two ends: BC, which is stable, and AD, which is mobile and tight fitting, and let AD be loaded from above with lead E which weighs 10,000,000 *libbre*.[8] I believe that you will understand how much violence the end BC might feel. Now if we force a plane or knife FG so that it enters and cuts the compressed wool, I say that if the wool $FBCG$ remains compressed like before, even if the end BC no longer feels any of the weight above from lead E, it still will experience the same weight that it experienced previously.[9] Try it yourself so that I won't bore you further.
>
> In regard to the second: There was once a philosopher who, seeing the spout put in a barrel by a servant, challenged him by saying that the wine would never come out, because the nature of weights is to press down and not horizontally or from every direction. But the servant proved to him that, even if liquids do gravitate downwards by nature, they push and splatter in every direction, even upwards, until they find places to arrive—*i.e.*, places that resist with a force less than the force of these liquids. Submerge a jug in water, with the mouth pointed downward, then make a hole in the bottom, so that the air can escape, and you will see with what impetus the water moves upwards from underneath to fill it. Try it yourself so that I won't bore you further.[10]

For my purposes here, the most interesting thing about Torricelli's response is that he is forced to clarify a more sophisticated version of his earlier claim that the weight of the air is the cause of the height to which the mercury is raised. The fact that Torricelli responds as he does show that he recognizes Ricci's causal-experimental arguments as valid, at least against the simplistic claim that the *weight* of the air (where weight is understood as a downward force) is the cause of the mercury rising. Torricelli may never have intended to defend that simplistic claim in the first letter, but after the misunderstanding, his second letter clarifies

[8] Martini (1883) gives the conversion for the *libbra* as .339542 kg in Florence and .339072 kg in Rome.

[9] Since Torricelli claims that the end BC no longer feels any of the weight from above, he must intend that the cutting plane FG rests on the exterior cylindrical vessel and supports the weight above.

[10] My translation. Original text of the letter can be found in Loria and Vassura (1919, 198–201).

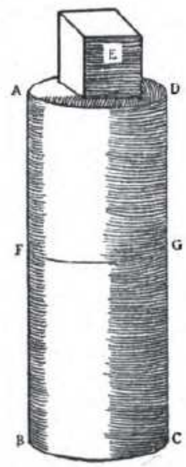


Figure 4.2: Diagram of Torricelli's thought experiment about compressible wool, included in his letter to Ricci, 28 June 1644. From Loria and Vassura (1919, 199).

his more sophisticated view about what is causing the mercury's rise: it may not be the downward weight of the atmosphere itself, but a general omni-directional compression of the air created by the entire mass of the downward weight of the atmosphere above.

In his response to Ricci's first causal-experimental argument, he uses the image of the wool to argue that, even if the weight of the air pressing down is completely supported by the rigid plane FG, the compression of the wool continues to exert exactly as much pressure as the downward weight had previously. By using this image, he undercuts the presumed difference between the two experimental instances (2A and 2B) that Ricci had provided. Although the basin in instance 2A was exposed to the atmosphere and the basin in 2B was sealed, Torricelli says that the compression of the air in the basin is identical in both cases—in 2A, it is compressed by the weight of the atmosphere, and in 2B, it is compressed by the rigid lid. On his theory, the height of the mercury would be expected to be identical in both cases.

In responding to Ricci's second causal-experimental argument, Torricelli explains that, just as liquids press in every direction (specifically, they "find places to arrive" that resist with a force less than their own), so also we should expect air to do this. Since we have the weight of an entire atmosphere of air pushing down on us, we should expect the air to push in every direction, just as we see happens at the bottom of any container of liquid. Thus, the contrast is undermined between Ricci's instances 3A and 3B. In 3A where the syringe is oriented horizontally, the air pushes with just as much force as it does in 3B where the plunger is oriented upward. The fact that there is equal resistance when we pull back the plunger in both cases is entirely to be expected, on Torricelli's theory.

Torricelli does not back down on his original claim that the vacuum is *not* a cause of the mercury rising, but he does explain a more sophisticated positive causal theory than was evident in his first letter. In general, some type of distributed force pushing on the surface of the liquid (due to the compression of the air) is the cause of its having a raised height in the tube.

4.2 ROBERVAL'S EXPERIMENTS

The Torricellian experiment soon became known in France after Ricci wrote about it to Marin Mersenne. French experimentalists were quickly taken by the experiment and began to perform a number of variations on Torricelli's experimental setup. Gilles Personne de Roberval performed a few interesting variations, of which I will here describe two.

Torricelli had arrived at the claim that the cause of mercury rising in his apparatus was an omni-directional compression of the air due to the weight of the atmosphere. In Torricelli's hands, however, this remained only a hypothesis; he had given only a negative experimental argument showing that the vacuum is not the cause of the mercury's height, not a positive experimental demonstration showing that altering the compression of the air might change the height of the mercury. Roberval was able to provide direct evidence of the air's capacity for expansion and compression and the force it exerts by virtue of this capacity.

Roberval performed a variation on Torricelli's experiment in which he added regular atmospheric air to the mercury in the tube before inverting it.[11] He allowed space for one

[11]The experiment is described in a 1648 letter from Roberval to Des Noyers, which was published in Pascal's works (Pascal 1908, 313–320). Webster (1965, 497–499) gives an excerpt in English translation. Roberval also describes a variation in which he added water as well as air to the tube, but he does not

and a half inches of regular air when filling the tube with mercury, and noticed that the resulting height of the mercury when he upended the tube in a vat of mercury was four inches less than the normal height when no air is admitted.[12] Thus, his experiment involved the following contrasting instances:

4A. The tube is entirely filled with mercury, stopped with a finger, and upended in a larger basin of mercury. The height of mercury in the tube falls to 2 and 7/24 feet.

4B. The tube is mostly filled with mercury, leaving one and a half inches of space filled with air at the top. The tube is then stopped with a finger and upended in a basin of mercury. The height of mercury in the tube falls four inches shorter than in the previous instance.[13]

Assuming identical background conditions and auxiliary interventions, these two instances constitute an experimental series for testing if the added air is a cause of the height of the mercury. The only difference between the two instances is the addition of 1-1/2 inches of atmospheric air to the tube, and so the 4-inch difference in the height of the mercury can only be attributed to the added air.

But what about the added air—what property—caused the difference in the height of the mercury? It could not be the *weight* of the added air that caused the difference. If we think of the Torricellian apparatus as a balance, with the weight of the atmosphere of air on one side and the weight of the column of mercury on the other, the weight of the added air on the side of the mercury could not have made up for the 4 inches of displaced mercury. Therefore, it had to be some *other* property of the added air which caused such a big change in the height of the mercury. For Roberval, the contrasting cases were an impressive proof of a property of air entirely independent of weight:

> The same air [trapped in the tube], while seeking to fill the whole space exerts a pressure in all directions, on the adjacent bodies, the tube keeps in the particles from all other directions, mercury being the only one of them that can give way towards the lower part

clearly describe his results. A similar experiment was performed around the same time by Ètienne Noël. See Webster (1965, 448).

[12] He calls it "our air", indicating that the one and a half inches were measured and admitted to the tube before upending the tube.

[13] "For, in that case a depression of the customary level by wholly four inches is seen, so that it did not ascend to the said height of two feet" (Webster 1965, 497).

> of the tube. Moreover this explanation must be adopted: that the air which we respire only possesses such force towards dilatation and rarefaction as is equal to the power of the natural element compressing or condensing it. Besides which, this also agrees with the laws of nature itself and applies in all other bodies which nature has granted powers of spring, as in bows and innumerable other examples. All of which bodies, as long as they are compressed by force, but not extended beyond the limits of their own power, never cease to exist. They are carried by an innate force of resilience, which is the same as that force with which they are drawn or impelled by other bodies.[14]

As we know, Torricelli had already hinted at a force of compression in using his analogy of compressed wool, but in this passage Roberval displays a more elaborate understanding of a kind of "spring" or "innate force of resilience" that is possessed by the air, independent of the force of its weight. And so Roberval constructed another more powerful experimental demonstration of this property of the air:

> I began to ponder in my mind, if perhaps there was some body available to us, which was both flexible and satisfactorily hold air. The convenient thing which came to my mind was the swim-bladder of the carp, because it is quite flexible and is thought to have been given to this animal by nature for the express purpose of containing air. Now this bladder has a double structure and the two parts are connected together by a narrow neck through which the air communicates. Of the two parts I selected the one which is more pointed and more nearly approaches the form of a cone, because the membrane of this second part is far stronger and splits with greater difficulty.
>
> This is now emptied of nearly all the air, so that the proportion of air remaining in it was not in fact 1000th part of that which it had formerly held. A thread was tied round the neck and I tied it so tight that it could not let out its air, nor admit any. This was then placed in the tube in which we had previously placed small birds and mice, the superior part of which has the capacity of a goose egg.[15]
>
> This being prepared, I made the experiment using mercury, so that the space or seeming vacuum appeared as usual at the upper part of the tube which held the bladder.[16] But to the complete astonishment of the bystanders, the bladder appeared quite turgid and distended, just as if it was still inside the carp's belly, for, in fact, that very small amount of air which remained in it, liberated at last from compression, being in a position in which it was no longer compressed neither by our condensed air, nor by other surrounding bodies, had expanded itself to the size which the bladder would permit [see figure 4.3]. And with the inclination of the tube, the mercury was sucked back, the bladder became flaccid, just as if its air was exhausted. Upon re-erecting the same tube, the mercury fell, the bladder expanded again.
>
> At length, by virtue of perforating the pig's bladder, which closed the upper end of the tube, using a fine needle, with but the minutest hole so that air gradually penetrated

[14] Excerpt of letter from Roberval to Pierre des Noyers, dated May 1648. The original letter is in Latin. English translation taken from Webster (1965, 498).

[15] Presumably, Roberval refers to experiments done to test the effect of the vacuum (or near vacuum) on on small animals.

[16] "The experiment using mercury" is a reference to the Torricellian apparatus.

the tube, and the air condensed around the bladder. The bladder deflated and gradually subsided until it returned to the state which it had been in when it was placed in the tube. Otherwise, at another time, if the hole was larger, the air rushed in, in a moment, the deflation would occur more rapidly.

All this confirmed our assertions about the air's rarefaction and condensation so that no one can any longer doubt it. ... This experiment had been tried more than a hundred times, in public and privately, in various ways and I had never failed. ...

I experimented also, in the same way, using a bladder full of our condensed air and properly tied up, and that air in the space or apparent vacuum being held back only by the membrane of the bladder, pressed by force on all sides, seeking to dilate, so that it burst some of them, principally around the thread which tied them, certainly because the membrane was weakened at that point.[17]

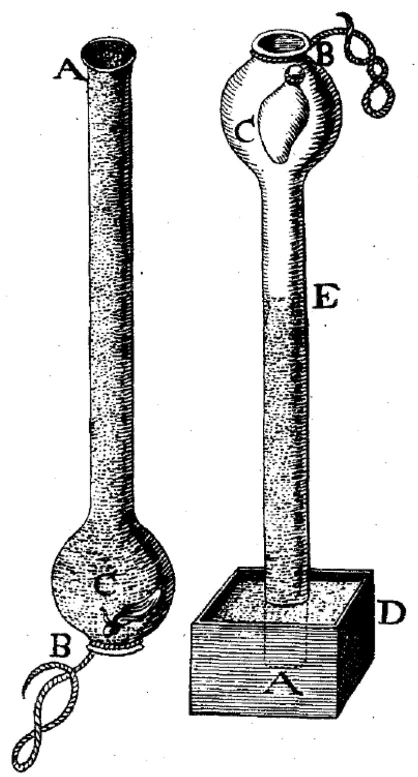


Figure 4.3: Diagram of Roberval's experiment with the carp bladder, 1648. From Pecquet (1653, 95).

Again, in this account, we see a clear use of contrasting experimental instances. The initial contrast which was so astonishing to his bystanders is the contrast between the bladder's flattened appearance in normal atmospheric air (*i.e.*, before the tube was filled with

[17]Excerpt from same letter as above. Translation again taken from Webster (1965, 496–497).

mercury) and its appearance once it was in the vacuum space of the apparatus after the mercury had fallen. We might describe the two instances as follows:

5A. The carp bladder is emptied of all air and a thread is tied tightly around its neck. The bladder is placed in the upper part of a Torricellian tube. A pig's bladder is used to seal the upper end of the tube, but the inside of the tube and the carp's bladder are still exposed to atmospheric air. The bladder appears flat and empty.

5B. The carp bladder is prepared and placed in the Torricellian tube as in 5A, with a pig's bladder sealing the upper end of the tube. The tube is turned upside-down and filled with mercury. The tube is then stopped with a finger and upended in a large basin of mercury. The mercury in the tube falls and reveals that the carp bladder now appears to be dilated as if full.

Note that these two instances do not technically satisfy the criteria for an experimental series, since the procedures for each instance differ by more than one intervention. For example, when confronted with these two instances, one could wonder if contact with mercury somehow altered the carp bladder in a way that changed its appearance. Perhaps because of this inadequacy in the above pair of instances, Roberval continued on in his account to provide several more satisfying contrasts. First, he provides a better contrasting instance to pair with instance 5B:

6A. Identical to 5B, with the end result of the carp bladder appearing dilated.

6B. Identical procedure to 6A, with the end result of the carp bladder appearing dilated. The tube is then tipped up so that the mercury pours out into the basin, and the bladder deflates again to a flat appearance.

Instances 6A and 6B now constitute an experimental series. Since the bladder in both instances has come into contact with the mercury, the only difference is in the final intervention of 6B which allows the mercury to pour out of the tube so that the bladder is re-exposed to atmospheric air. The contrast of these two instances supports the claim that the exposure of the bladder to atmospheric air (as opposed to its exposure to the empty space of the Torricellian apparatus) is a cause of the bladder's flattened and empty appearance. The same bladder, voided as much as possible of internal air in both cases, displays a striking difference in volume between the ordinary circumstances of our "condensed air" and its circumstances inside the Torricellian vacuum.

Roberval's account of making a tiny hole in the pig bladder can likewise be read as a continuum of experimental instances that constitutes one experimental series.

7A. Identical to 5B and 6A, with the end result of the carp bladder appearing dilated.

7B. Identical to 7A, with the end result of the carp bladder appearing dilated. Using a pin, make a tiny hole in the pig bladder that is fitted over the upper end of the tube. After one small unit of time (τ) observe that the carp bladder is slightly less dilated than in 7A.

7C. Identical to 7B. After two small units of time (2τ) observe that the carp bladder is slightly less dilated than in 7B.

$\vdots$

7Z. Identical to 7B. After n small units of time ($n\tau$) observe that the carp bladder is completely flattened in appearance.

The only difference among the above instances is the amount of air that had been allowed to leak from the atmosphere into the enclosed space in the apparatus. The more air that entered, the more compressed the carp bladder became.

As Roberval notes at the end of the above passage, he also tried a contrast between the two instances like 5A and 5B, but starting with a bladder full of atmospheric air instead of a voided one. All of his experimental series show that the very same quantity of air enclosed in a bladder (whether initially a minuscule amount amount or a larger amount), when placed under different external compression by intervention, experiences drastic changes in volume. Thus, on the basis of Roberval's experiments and his theoretical understanding of it, he could justifiably make the claim that the quantity of external compression is a cause of the volume of a contained quantity of air. However, he did not yet have any quantitative measure of the causal relationship.

4.3 PASCAL'S EXPERIMENTS

Blaise Pascal knew of Torricelli's experiment by way of Mersenne and devised a variation of his own, which he called the "vacuum within a vacuum" experiment. He described the experiment in a letter to Florin Pèrier in 1647:

> [I] recall to you the experiment I made lately in your presence with the two tubes, one inside the other, which exhibit a vacuum within a vacuum. You saw that the quicksilver of the inner tube hung suspended at the usual height when it was counterpoised and pressed by the weight of the whole mass of the air, but that it dropped altogether, so that it was no longer suspended at all, when by removing all the surrounding air we made a complete vacuum about it so that it was no longer pressed in by it and counterbalanced. Afterward, you saw that the height or suspension of the quicksilver increased or decreased as the pressure of the air increased or decreased, and finally that all these various heights or suspensions of the quicksilver were always proportional to the pressure of the air.[18]

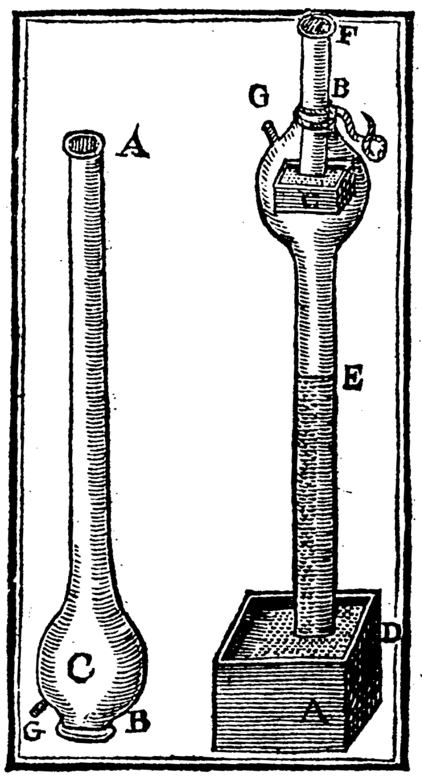


Figure 4.4: Illustration of one type of vacuum-in-a-vacuum experiment, likely similar to Pascal's setup. Taken from a description of Auzout's experiment in Pecquet (1653, 105).

[18]The letter is dated November 15, 1647. This translation is from Pascal (1937, 99–100).

What Pascal had essentially done was to put one Torricellian apparatus within another Torricellian apparatus, to see what the height of mercury in the former apparatus would be when not in contact with atmospheric air.[19] The experiment displays a contrast between a Torricellian apparatus whose vessel of mercury is exposed to the atmosphere, and a Torricellian apparatus whose vessel of mercury is exposed to a vacuum:

8A. Place vessel AB upright in a large container D of mercury. Insert a small square container in the globe of vessel AB and insert pipe CF so that it is standing upright through end B. Seal end B around pipe using a pig bladder, and also seal access hole G. Keeping the entire apparatus erect, stop the bottom end A with a finger and pour mercury into the top of the upper pipe F until the entire vessel AB and pipe F are full of mercury. Seal entrance F with a pig's bladder. Remove finger from end A and observe that the mercury in vessel AB falls to a height of 27 inches, and the mercury in the small tube CF falls to the same level as that in the small square container inside the globe of vessel AB.

8B. Identical to 8A, with end result of mercury in tube CF being at same level as that in square container inside globe. Then poke a small hole in the bladder sealing access hole G and observe that as the air slowly re-enters the globe of vessel AB, the mercury in pipe CF slowly rises until it ultimately stands at a height of 27 inches.

Both of these experimental instances can be performed under similar background conditions, and the procedure of auxiliary interventions is identical in both cases except for the final intervention in 8B in which a small hole is made in the bladder covering access hole G. The observed height of mercury in the tube CF differs in the two cases, and so the pair of instances is evidence for the claim that atmospheric pressure is the cause of the height of mercury in the tube. We see in the quotation above from Pascal that he expresses a clear understanding here that the height of the mercury is proportional to the pressure of the air.

However, although Pascal was inclined to believe that the pressure of the air was the true cause of the elevation of the mercury, he was not ready, neither on the basis of the classic Torricellian experiment nor on the basis of his own vacuum-within-a-vacuum experiment, to

[19] A similar experiment was performed around the same time by Adrien Auzout and is described by Pecquet (1653, 104–109).

dismiss the traditional hypothesis of nature's abhorrence of a vacuum. Here are his comments on the matter:

> Opinions have been divided: some have been content to say only that nature abhors a vacuum, others have maintained that she could not tolerate it. I have tried in my pamphlet on the vacuum to refute the latter opinion, and I believe that the experiments recorded there suffice to show indubitably that nature can, and does, tolerate any amount of space empty of any of the substances that we are acquainted with, and that are perceptible to our senses. I am now engaged in testing the truth of the former statement, namely, that nature abhors a vacuum, and am trying to find experimental ways to show whether the effects ascribed to the abhorrence of a vacuum are really attributable to that abhorrence, or to the weight and pressure of the air. For, to reveal to you frankly my whole thought on the matter, I can hardly admit that nature, which is not at all animated or sensible, can be capable of abhorrence, since the passions presuppose a soul capable of experiencing them. I feel much more inclined to attribute all these effects to the weight and pressure of the air, because I consider them only as particular cases of a universal principle concerning the Equilibrium of Fluids, which is to be the greater part of the Treatise I have promised. Not but that I had the same thoughts when I brought out my abridgment; but for lack of convincing experiments, I dared not then (and I dare not yet) give up the idea of the abhorrence of a vacuum. . . . Indeed I do not consider that it is permissible for us lightly to discard the maxims handed down to us by the ancients unless we are compelled to do so by indubitable and unanswerable proofs. . . .
>
> After that experiment [the void in a void experiment] *we certainly had reason to believe*, as we do believe, that it is not the abhorrence of the vacuum that causes the quicksilver to stand suspended in the usual experiment, but really the weight and pressure of the air, which balances the weight of the quicksilver. *But seeing that all the effects of this last experiment with the two tubes, which is so naturally explained by the mere pressure and weight of the air, can also be explained, probably enough, by the abhorrence of a vacuum*, I still hold to that ancient principle, although I am determined to seek a thorough elucidation of this difficulty by means of a decisive experiment.[20]

Pascal took the Torricellian experiment and his own vacuum-within-a-vacuum experiment to show that a vacuum can indeed be achieved. The question of whether or not that vacuum exerted some force in virtue of nature's abhorrence of a vacuum, however, he took to be open to debate. He hesitated to say that the experiments were decisive evidence for refuting the view that nature's abhorrence of a vacuum is the cause of elevation of the mercury.

Why might he have thought this, when Torricelli thought that his own experimental-causal argument had refuted the traditional view?[21] The difference between Torricelli and

[20] Emphasis mine. Passage is from same letter as above; translation taken from Pascal (1937, 99–100).

[21] See section 4.1.

Pascal on this point most likely comes down to a difference in opinion about how the force of the vacuum might work. For Torricelli, a difference in the size of the space above the mercury was understood to coincide with a difference in any theoretical force that a vacuum might be able to exert on the mercury. Perhaps Pascal, in contrast, did not assume any connection between the size of the volume occupied by the apparent vacuum and the force it could exert. For a believer in the force of the vacuum who does not hold the distinct belief that a different amount of force is exerted by different spatial volumes occupied by that vacuum, the Torricellian experiment cannot prove that the vacuum is not a cause of the elevation of the mercury.

Evidently, Pascal also did not consider his vacuum-in-a-vacuum experiment to be definitive proof of the claim that atmospheric pressure causes the height of the mercury. In the passage above, he says that the principle of abhorrence of a vacuum can explain the phenomena of his experiment. A firm believer in the principle of nature's abhorrence to a vacuum might explain the observation in instance 8A by saying that the mercury is pulled equally by the force of the vacuum on both sides (both from the vacuum in the tube CF and from the vacuum in the globe of vessel AB). Instance 8B can likewise be explained on the same theory: the hole poked in the entryway G allows air to enter and the force of the vacuum within the globe is decreased, so the mercury in tube CF rises, since the force of the vacuum within CF remains as strong as ever.

It is worth emphasizing that Pascal remains uncertain of the causal claim *not* because he rejects the use of contrasting experimental instances for testing causal claims, but because he is uncertain of *what exactly* differed between the two contrasting instances in his experiment. His uncertainty lies in determining what the final intervention in instance 8B achieved: did poking the hole in entryway G reduce the force of the vacuum, or did it rather increase the air pressure?

We might wonder what *would* constitute experimental evidence for Pascal that it is the weight and pressure of the air, rather than the force of the vacuum, that causes the elevation of mercury in the Torricellian apparatus. Fortunately, he tells. In the letter to Pèrier, he continues on to describe just such an experiment:

> To this end I have devised one [experiment] that is in itself sufficient to give us the light we

> seek if it can be carried out with accuracy. This is to perform the usual experiment with a vacuum several times over in one day, with the same tube and with the same quicksilver, sometimes at the base and sometimes at the summit of a mountain at least five or six hundred fathoms high, in order to ascertain whether the height of the quicksilver suspended in the tube will be the same or different in the two situations. You see at once, doubtless, that such an experiment is decisive. If it happens that the height of the quicksilver is less at the top than at the base of the mountain (as I have many reasons to believe it is, although all who have studied the matter are of the opposite opinion), it follows of necessity that the weight and pressure of the air is the sole cause of this suspension of the quicksilver, and not the abhorrence of a vacuum: for it is quite certain that there is much more air that presses on the foot of the mountain than there is on its summit, and one cannot well say that nature abhors a vacuum more at the foot of the mountain than at its summit (Pascal 1937, 100–101).

Notice the explicit experimental-causal reasoning taking place in the above passage. Pascal endeavors to create experimental instances that are entirely identical—even in regard to the force potentially exerted by nature in its "abhorrence of a vacuum"—*except* for the quantity of air pressing down on the apparatus. If, then, a difference is seen in the height of the mercury, it can only be attributed to the weight and pressure of the air, and not to any force exerted by the vacuum.

For lack of high hills or mountains surrounding Paris, Pascal asked Pèrier to perform the experiment for him on Puy de Dôme, which neighbors the town of Clermont where Pèrier was living. Pèrier did perform the experiment, which he described in a return letter almost a year later:

> On that day, therefore, at eight o'clock in the morning, we started off all together for the garden of the Minim Fathers, which is almost the lowest spot in the town, and there began the experiment in this manner.
>
> First, I poured into a vessel six pounds of quicksilver which I had rectified during the three days preceding; and having taken glass tubes of the same size, each four feet long and hermetically sealed at one end but open at the other, I placed them in the same vessel and carried out with each of them the usual vacuum experiment. Then, having set them up side by side without lifting them out of the vessel, I found that the quicksilver left in each of them stood at the same level, which was twenty-six inches and three and a half lines above the surface of the quicksilver in the vessel.[22] I repeated this experiment twice at this same spot, in the same tubes, with the same quicksilver, and in the same vessel; and found in each case that the quicksilver in the two tubes stood at the same horizontal level, and at the same height as in the first trial.
>
> That done, I fixed one of the tubes permanently in its vessel for continuous experiment. I marked on the glass the height of the quicksilver, and leaving that tube where it stood,

[22]Each line marks one twelfth of an inch, or 0.212 cm.

> I requested Revd. Father Chastin, one of the brothers of the house, a man as pious as he is capable, and one who reasons very well upon these matters, to be so good as to observe from time to time all day any changes that might occur. With the other tube and a portion of the same quicksilver, I then proceeded with all these gentlemen to the top of the Puy de Dôme, some 500 fathoms above the Convent. There, after I had made the same experiments in the same way that I had made them at the Minims, we found that there remained in the tube a height of only twenty-three inches and two lines of quicksilver; whereas in the same tube, at the Minims we had found a height of twenty-six inches and three and a half lines. Thus between the heights of the quicksilver in the two experiments there proved to be a difference of three inches one line and a half. We were so carried away with wonder and delight, and our surprise was so great that we wished, for our own satisfaction, to repeat the experiment. So I carried it out with the greatest care five times more at different points on the summit of the mountain, once in the shelter of the little chapel that stands there, once in the open, once shielded from the wind, once in the wind, once in fine weather, once in the rain and fog which visited us occasionally. Each time I most carefully rid the tube of air; and in all these experiments we invariably found the same height of quicksilver. This was twenty-three inches, one line and half in comparison with the twenty-six inches, three lines and a half which had been found at the Minims. This satisfied us fully.[23]

Pèrier continued on in the letter to explain that, on his descent back down the mountain, he made more measurements at an intermediate point, finding a height of twenty-five inches of mercury. He also reported that Father Chastin confirmed that the apparatus at the bottom remained at the same level of mercury all day. He concluded the letter with a report of measurements taken at various altitudes, including the highest tower of the town and an additional point on the mountain. As an experiment that provides evidence for a causal claim, it can be divided into the following instances:

9A. Set up the Torricellian apparatus at the lowest spot in town. Observe that the mercury stands at a height of 26 inches + 3.5 lines. (It remains at this level all day).

9B. Set up the Torricellian apparatus at the summit of Puy de Dôme. Observe that the mercury stands at a height of 23 inches + 2 lines.

9C. Set up the Torricellian apparatus at an intermediate point on the mountain. Observe that the mercury stands at a height of 25 inches.

For this series to be a sufficient test of the claim that the pressure of the air is the cause of the height of the mercury, these instances must be identical in all ways except the altitude at which they are performed (and by hypothesis, the air pressure varies with altitude).

[23]This letter is dated September 22, 1648. Translation from Pascal (1937, 104–105).

In the passage, Pèrier has an obvious concern with establishing that the conditions of the experimental instances were sufficiently similar. He spends time explaining that he used identical tubes and the same mercury in each case, and even performs a control instance of the experiment at the low point to make sure that two separate instances of the apparatus give the same result. The many repetitions of the experiment on the summit of the mountain are important to provide evidence that the differences in the height of the mercury between circumstances of shelter vs. open air, wind vs. no wind, and fine weather vs. rain and fog are negligible. (Essentially these instances provide evidence, but not proof, that each of the differences of circumstances is *not* an appreciable cause of the height of the mercury.) In addition, the stationing of Father Chastin with the apparatus at the low point serves to give evidence that time of day or differing weather conditions might be relevant to the way that the result differs with altitude.

Pascal, satisfied with this result, wrote: "I now find no difficulty in accepting . . . that nature has no repugnance to a vacuum, and makes no effort to avoid it; that all the effects ascribed to such abhorrence are due to the weight and pressure of the air, which is their only real cause".[24]

4.4 BOYLE'S EXPERIMENTS AND THEORIZING

Accounts of experiments done with the Torricellian apparatus began to reach England by means of correspondence and popular publications. Robert Boyle had heard of these experiments and may have reproduced some of them himself, but he only began his first concentrated study of air pressure in 1658 once he had an air pump. The pump had been invented by Otto von Guericke in 1654, and Boyle had heard of it and had one constructed for himself.[25] Boyle investigated many phenomena in vacuum—*e.g.*, the expansion of bladders, the extinguishing of flames, and magnetic attraction. His primary goal, however, was to try

[24] Pascal published these comments along with the above-cited letters in a 1648 printed account entitled "Story of the Great Experiment on the Equilibrium of Fluids". Translation taken from (Pascal 1937, 110).

[25] Robert Hooke, who served as Boyle's assistant at Oxford, designed and constructed the pump with some assistance from the London instrument maker Ralph Greatorex. Boyle heard of the pump through the account of it in Gaspar Schott's 1657 *Mechanica Hydraulico-Pneumatica*.

the vacuum-within-a-vacuum experiment by replacing the outer Torricellian apparatus with the vacuum chamber of the air pump.[26] He had hopes that the experiment would allow him "to give a near guess at the proportion of force betwixt the pressure of the Air (according to its various states, as to Density and Rarefaction) and the gravity of Quick-silver, then hitherto has been done" (Boyle 1660, 115). Due in part to difficulties in carrying out the experiment (especially leaking of the vacuum chamber, difficulty reading the exact level of the mercury) and also to mathematical difficulties in obtaining the exact volume of the air in the vacuum chamber at any given moment,[27] he was unable to arrive at a mathematical expression of the relationship between the density or pressure of the air and the height of the mercury. Thus, he had to devise a different experiment.

4.4.1 Boyle's J-tube experiment

In the 1662 edition of *New Experiments Physico-Mechanical*, however, Boyle described two new experiments in which he was able to provide evidence of a specific mathematical relationship.[28] The first of these was an experiment with compressed air in a J-shaped tube:

> We took a long Glass-Tube, which by a dexterous hand and the help of Lamp was in such a manner crooked at the bottom, that the part turned up was almost parallel to the rest of the Tube, and the Orifice of this shorter leg of the Siphon (if I may so call the whole Instrument) being Hermetically seal'd, the length of it was divided into Inches, (each of which was subdivided into eight parts) by a straight list of paper, which containing those Divisions was carefully pasted all along it: then putting in as much Quicksilver as served to fill the Arch or bended part of the Siphon, that the *Mercury* standing in a level might reach in the one leg to the bottom of the divided paper, and just to the same height or Horizontal line in the other; we took care, by frequently inclining the Tube, so that the Air might freely passfrom [sic] one leg into the other by the sides of the *Mercury*, (we took (I say) care) that the Air at last included in the shorter Cylinder should be of the same laxity with the rest of the Air about it. This done, we began to pour Quicksilver into the longer leg of the Siphon, which by its weight pressing up that in the shorter leg, did by degrees streighten the included Air: and continuing this pouring in of Quicksilver till the

[26]This experiment is described as Experiment 17 in the 1660 edition of *New Experiments Physico-Mechanicall*. There, Boyle refers to it as "that Experiment, whereof the satisfactory tryal was the principal Fruit I promis'd my self from our Engine" (Boyle 1660, 106).

[27]Giving an exact measure for the volume would involve subtracting the volume of the apparatus and also accounting for the changing level of mercury in the reservoir. Boyle's account of the experiment mentions all of the practical difficulties he encountered and some of the mathematical considerations which would need to be taken into account. See especially Boyle (1660, 113–118).

[28]The new experiments are described in a separate appended section entitled *A Defence of the doctrine touching the spring and weight of the air*, which formed a defense of his views agains Linus.

> Air in the shorter leg was by condensation reduced to take up but half the space it possess'd (I say, *possess'd*, not *fill'd*) before; we cast our eyes upon the longer leg of the Glass, on which was likewise pasted a list of Paper carefully divided into Inches and parts, and we observed, not without delight and satisfaction, that the Quicksilver in that longer part of the Tube was 29 Inches higher then the other. Now that this Observation does both very well agree with and confirm our *Hypothesis*, will be easily discerned by him that takes notice that we teach, and Monsieur *Paschall* and our *English* friends [sic] Experiments prove,[29] that the greater the weight that leans upon the Air, the more forcible is its endeavour of dilatation and consequently its power of resistance, (as other Springs are stronger when bent by greater weights.) For this being considered, it will appear to agree rarely well with the *Hypothesis*, that as according to the Air in that degree of density and correspondent measure of resistance to which the weight of the incumbent Atmosphere had brought it, was able to counterbalance and resist the pressure of a Mercurial Cylinder of about 29 Inches, as we are taught by the *Torricellian* Experiment; so here the same Air being brought to a degree of density about twice as great as that it had before, obtains a Spring twice as strong as formerly.

After obtaining these preliminary results showing that air with twice the density supports twice the external pressure, the J-shaped tube broke. After constructing another longer tube, Boyle gathered more extensive results of how much mercury would be supported by differing compressed volumes of air (see figure 4.5).

Each line in Boyle's table in figure 4.5 refers to a single experimental instance, so that there are a total of 25 instances. All of these instances together form an experimental series for testing if the quantity of mercury pressing on an enclosed volume of air is a cause of the volume occupied by that air. The instances were all performed under similar background conditions using an identical procedure except for the final intervention that varied the amount of mercury poured into the apparatus. By varying the quantity of mercury while keeping all else the same and showing that the volume of the air changes, the experimental series is excellent evidence that the external pressure on an enclosed quantity of air is a cause of its volume. In contrast to the experimental series that have been discussed in the preceding sections, however, this is the first experimental series that supports a specific *quantitative* relationship between a cause and effect.

Column E in Boyle's table (see figure 4.5) gives calculated pressures based on the hypothesis "that supposes the pressures and expansions to be in reciprocal proportion." In

[29] The English experiments to which Boyle refers are likely the experiments of Townley and Power, which were performed in the spring of 1661 and reported in Power (1664, 127). See Webster (1965) for a discussion of their experiments.

A Table of the Condensation of the Air.

A	*A*	*B*	*C*	*D*	*E*
48	12	00	Added to $29\frac{1}{8}$ makes	$29\frac{2}{16}$	$29\frac{2}{16}$
46	$11\frac{1}{2}$	$01\frac{7}{16}$		$30\frac{9}{16}$	$30\frac{6}{16}$
44	11	$02\frac{13}{16}$		$31\frac{15}{16}$	$31\frac{12}{16}$
42	$10\frac{1}{2}$	$04\frac{6}{16}$		$33\frac{8}{16}$	$33\frac{1}{7}$
40	10	$06\frac{3}{16}$		$35\frac{5}{16}$	35 --
38	$9\frac{1}{2}$	$07\frac{14}{16}$		37 --	$36\frac{15}{19}$
36	9	$10\frac{2}{16}$		39 [illegible]/16	$38\frac{7}{8}$
34	$8\frac{1}{2}$	$12\frac{8}{16}$		$41\frac{10}{16}$	$41\frac{2}{17}$
32	8	$15\frac{1}{16}$		$44\frac{3}{16}$	$43\frac{11}{16}$
30	$7\frac{1}{2}$	$17\frac{15}{16}$		$47\frac{1}{16}$	$46\frac{3}{5}$
28	7	$21\frac{3}{16}$		$50\frac{5}{16}$	50 --
26	$6\frac{1}{2}$	$25\frac{3}{16}$		$54\frac{5}{16}$	$53\frac{10}{13}$
24	6	$29\frac{11}{16}$		$58\frac{13}{16}$	$58\frac{2}{8}$
23	$5\frac{3}{4}$	$32\frac{3}{1}$		$61\frac{5}{16}$	$60\frac{18}{23}$
22	$5\frac{1}{2}$	$34\frac{15}{16}$		$64\frac{1}{16}$	$63\frac{6}{11}$
21	$5\frac{1}{4}$	$37\frac{15}{16}$		$67\frac{1}{16}$	$66\frac{4}{7}$
20	5	$41\frac{9}{16}$		$70\frac{11}{16}$	70 --
19	$4\frac{3}{4}$	45 --		$74\frac{2}{16}$	$73\frac{11}{19}$
18	$4\frac{1}{2}$	$48\frac{12}{16}$		$77\frac{14}{16}$	$77\frac{2}{3}$
17	$4\frac{1}{4}$	$53\frac{11}{16}$		$82\frac{12}{16}$	$82\frac{4}{17}$
16	4	$58\frac{2}{16}$		$87\frac{14}{16}$	$87\frac{3}{8}$
15	$3\frac{3}{4}$	$63\frac{15}{16}$		$93\frac{1}{16}$	$93\frac{1}{5}$
14	$3\frac{1}{2}$	$71\frac{5}{16}$		$100\frac{7}{16}$	$99\frac{6}{7}$
13	$3\frac{1}{4}$	$78\frac{11}{16}$		$107\frac{13}{16}$	$107\frac{7}{13}$
12	3	$88\frac{7}{16}$		$117\frac{9}{16}$	$116\frac{4}{8}$

AA. The number of equal spaces in the shorter leg, that contained the same parcel of Air diversly extended.

B. The height of the Mercurial Cylinder in the longer leg, that compress'd the Air into those dimensions.

C. The height of a Mercurial Cylinder that counterbalanc'd the pressure of the Atmosphere.

D. The Aggregate of the two last Columns *B* and *C*, exhibiting the pressure sustained by the included Air.

E. What that pressure should be according to the *Hypothesis*, that supposes the pressures and expansions to be in reciprocal proportion.

Figure 4.5: Boyle's table giving the volume of enclosed air and the height of mercury supported in the J-tube experiment (Boyle 1662, 60). The second column (to which I will refer as A_2) is measured in the inches of the actual experiment, while the first column (A_1) is simply a conversion to quarter-inch units. Column B is measured in inches of mercury, and column C reflects that the pressure of the atmospheric air was measured to be 29-1/8 inches at the time of the experiment. Column D gives the calculated total pressure on the enclosed air at the condensed volume, and column E gives the prediction of Boyle's hypothesis that the external pressure on the condensed air (D) is inversely proportional to its volume (which is proportional to A) by some expression of the form $A \times D = \alpha$.

other words, Boyle's hypothesis, in mathematical form, is that the external pressure (P) on a quantity of air and the volume (V) of that quantity of air are inversely proportional to one another:

$$P \times V = \alpha \tag{4.1}$$

where α is a constant across all of the experimental instances, or equivalently:

$$P_1 \times V_1 = P_2 \times V_2 \tag{4.2}$$

where P_1 and V_1 refer to the measures of pressure and volume of the gas at any one instance, and P_2 and V_2 refer to those measures at any other instance.

Before moving on to Boyle's second experiment, it will be advantageous to pause and analyze two distinct abstractive moves that have already taken place in Boyle's theorizing on the basis of this experiment. His first abstractive move was from several concrete (historical) experimental instances to a table of numerical results, and his second abstractive move was from those numerical results to a hypothesis about inverse proportionality.

4.4.2 First-order abstraction

What exactly happened in the first abstractive move? It is obvious—perhaps so obvious that we fail to notice—that there is a great disparity between an experiment and the data table or chart that an experimenter uses to represent the experiment. An experiment is a historical event, performed by a person or persons, at some location and point in time. A data table, on the other hand, is a numerical representation or summary of that event. The disparity between the experiment itself and the data table used to represent it is a point emphasized by Pierre Duhem. Duhem draws a distinction between the *concrete* or *practical* facts of an experiment—*i.e.*, the detailed, historical sequence of physical operations and events that make up the experiment—and the *theoretical* facts—*i.e.*, the quantitative representation of the concrete facts.

> Between the concrete facts, as the physicist observes them, and the numerical symbols by which these facts are represented in the calculations of the theorist, there is an extremely great difference (Duhem [1914] 1954, 133).

In order to begin working with concrete facts, the physicist must choose which aspects of those facts to attend to. Once he has done this, he must establish a quantitative scale (or refer to a pre-established one) in order to represent these aspects in a numerical form, and he must choose procedures and instruments for measuring them. The instruments and procedures themselves are designed and interpreted by theories that are already established (a process which Duhem refers to as *translation*). The numbers that result from the entire interpretative procedure, then, are called theoretical facts.

Duhem emphasizes that the process of translation from concrete to theoretical facts is a "treacherous" one (Duhem [1914] 1954, 133). One reason he says this is that a single practical fact can potentially be represented by an infinity of distinct theoretical facts, either because of the distinct choices in how to read a measurement instrument or in the involvement of theories that must be invoked in order to use measuring instruments and procedures in the first place. Here, I will not be concentrating on this issue of underdetermination, but rather on another treacherous feature of the translation: information loss.

In the process of abstracting from a concrete experiment to numerical data, information is always lost. Boyle actually performed the set-up of the experiment and the operations that he described, stopping to record measurements at various points as he poured more mercury into the tube. The historical sequence of events that constituted his enacting the experiment were filled with all sorts of details to which we will never be privy. These details range from the banal and irrelevant to items of possible importance. Was Boyle wearing yellow breeches with pink polkadots? Was it raining at the time? How warm was it in the room and was there any change over the course of the experiment? All of this information is left out of Boyle's report. In reporting certain facts and not others, he made implicit assumptions about the importance and relevance of *those* facts relative to these other possible facts just mentioned. Even those facts which Boyle did deem important enough to report in his written account of the experiment are *distilled down* into a chart of numbers.

Unfortunately, there is one particular kind of information that is nearly always left out of the mathematical representation of an experiment. Normally, there is no information in an experimental data table about what factors were intervened upon by the experimenter and what factors were merely observed. No distinction is made between an action performed to *make* a certain quantitative property take on a certain value, and the mere observation of a quantitative property that happened to have a certain value under the circumstances of the experiment; in a data table, both kinds of quantitative properties are merely represented by numerical measurements and are indistinguishable.

Recall, for example, that in the J-tube experiment, Boyle poured mercury into the long leg of the tube to increase the external pressure on the trapped air. In doing so, he intervened to make the external pressure on the trapped air take on a certain value. In contrast, the volume of the trapped air was observed as a response to the action of increasing the external pressure. Thus external pressure was intervened upon—*i.e.*, brought to a certain value—while volume was merely observed. Yet this distinction is not noted in Boyle's table in figure 4.5.

But as I have been emphasizing in previous chapters, the distinction between a factor set by intervention and a factor that is merely observed is crucial to the question of whether or not an experiment provides evidence of a causal relationship. If a factor A is genuinely a cause of another factor B, and only if this is the case, we expect that there will be some particular way of intervening on A such that B will also vary. It is information about intervention and a system's response to intervention that gives us causal information.

Therefore, a data table which provides no information about which factors have been intervened upon and which have been observed can provide evidence only for statistical associations, not for causal relationships. For example, from Boyle's data table for the J-tube experiment, we only know that 12 inches of trapped air *corresponded* to a total external pressure equivalent to the weight of $29\frac{2}{16}$ inches of mercury, and that 6 inches of trapped air *corresponded* to a total external pressure equivalent to $58\frac{13}{16}$ inches of mercury. The data table alone gives no evidence that one of these factors might be a *cause* of the other. Still, even though the data table only offers us information about a statistical correlation, the experiment itself was able to offer us evidence *beyond* just this correlation. We know from Boyle's own written account that he intervened on the external pressure by pouring more mercury into the tube. And from that information—*i.e.*, that when intervening on external pressure he saw a change in volume—we have evidence that external pressure is an interventionist cause of volume. Unfortunately, this causal information is something that becomes hidden when the results are represented in the data table; the information is hidden from someone who considers only the data table as an informative representation of the facts of the experiment. As I will argue below, this information becomes *latent content* in the physical theory for which the data table is evidence.

4.4.3 Second-order abstraction

The move from data tables to a physical theory is what I refer to as *second-order abstraction*. To begin discussing this move, we might start with a very basic question: Why do we seek to go beyond the experimental results represented in a data table? Why aren't they enough? What do we gain when we formulate a physical theory?

The simple answer is that, if we *don't* amplify, or go beyond, our experimental results, all we have is a natural history: an experiment was performed by such-and-such a person on such-and-such a date under such-and-such conditions, with certain measurements resulting. This kind of information isn't enough for us, because it remains at the level of *actual* fact. Actual fact, when considered alone, isn't all that interesting from the point of view of physical science.

Physical science is an attempt to explore the *possibilities* of the physical world. Not just what has actually happened on certain occasions, but what *could have happened* on those past occasions, and what *could happen* in the future. In fact, physical science seeks to go beyond even just past and future possibilities. It is a subject of theoretical interest for the physical scientist to discuss various possibilities *even if they never have been and never will be instantiated in our actual world history*. (Consider, for example, the theoretical interest of violations of the second law of thermodynamics which are technically possible, even though we never observe them and they may never *actually* occur.)

The fact that the physical sciences are really after *possibility* rather than *actuality* drives us in the attempt to amplify our theory beyond our actual experimental observations. Actual historical observations are a rather impoverished source in comparison to the possibilities that we might like to explore. There are several ways in which we ampliatively infer beyond actual results in order to theorize about physical possibilities: (1) interpolation and extrapolation,

(2) ampliative inference about background conditions, and (3) ampliative inference about the results of interventions.

4.4.3.1 Interpolation and extrapolation. In physical theorizing, we try to extend the tested range of values of the variables under consideration in an experiment or set of experiments. Take Boyle's J-tube experiment as an example. Boyle only tested the pressure range from normal atmospheric pressure up to four times atmospheric pressure. We know that he recorded results at both extremes and at several points in between. But naturally, we want to ampliatively infer to interpolated points between those recorded results. Even though Boyle didn't record in his chart, for example, what volume corresponded to 27 inches of mercury in the longer leg of the tube, we can calculate what the approximate results would likely have been. Boyle noticed that the data corresponded roughly to a reciprocal relationship between pressure and volume—a relationship which allowed for calculation of the expected values he gave in column E (see figure 4.5), and which would also allow for calculation of interpolated points.

In many cases, interpolation is a safe move. In the case of the J-tube experiment, for example, in order for Boyle to get from the recorded instance in which there were $25\frac{3}{16}$ inches of mercury to the state where there were $29\frac{11}{16}$ inches of mercury, there must have been a moment in which there were actually 27 inches, and presumably there was at that moment no sudden and unexpected spike or dip in the volume of trapped air.

Boyle's "hypothesis" (as he called it) that the pressures and volumes stand in reciprocal proportion also allows for extrapolation beyond the tested range of variable values. Extrapolation is usually more tenuous than interpolation, because we don't know the extent to which a relationship will hold and whether or not it will break down under more extreme variable values. Boyle himself was aware of this danger and explicitly commented on it in his reflections following his description of the J-tube experiment:

> Till further tryal hath more clearly informed me, I shall not venture to determine whether or no the intimated Theory will hold universally and precisely, either in Condensation of Air, or Rarefaction: all that I shall now urge being, That however, the tryal already made sufficiently proves the main thing for which I here alledge it; since by it 'tis evident, that as common Air when reduc'd to half its wonted extent, obtained near about twice as forcible a Spring as it had before; so this thus-compress Air being further thrust into half this narrow

> room, obtained thereby a Spring about as strong again as that it last had, and consequently four times as strong as that of the common Air. And there is no cause to doubt, that if we had been here furnisht with a greater quantity of Quicksilver and a very strong Tube, we might by a further compression of the included Air have made it counterbalance the pressure of a far taller and heavier Cylinder of Mercury. For no man perhaps yet knows how near to an infinite compression the Air may be capable of, if the compressing force be competently increast (Boyle 1662, 62).

Thus, Boyle states his confidence that higher compressions will correspond to higher columns of mercury, but is hesitant to claim that the specific pattern of inverse proportionality will hold up at very high compressions. It is also notable that Boyle lists condensation and rarefaction of air as two separate phenomena, making clear that he does not assume that extrapolation to pressures *less* than normal atmospheric pressure will continue to manifest the same relationship as pressures greater than atmospheric pressure. Whether or not the relationship would continue to hold at lower pressures was an empirical question that would have to be tested, and Boyle did exactly that in another experiment that I will discuss below.[30] For now, the important point is that physical theorizing—which necessarily involves abstracting beyond experiment and also beyond any data table summarizing the actual results of experiment—will require interpolation and extrapolation beyond the variable values actually tested by finite acts of experimentation.

4.4.3.2 Background conditions. A second way in which we ampliatively infer beyond the actual results of an experiment is by presuming that our particular measurements would still have held up had some background conditions been different. Under the category of "background conditions", I include anything that is not explicitly represented as a column in the data table. We know by common sense that some of the details of how the experiment was performed were unimportant to the result. We may presume that Boyle was not, in fact, wearing yellow breeches with pink polka-dots, but it also is blatantly obvious to us that his measurements would not have been different if he had chosen to dress differently on the day of the experiment. The clothing choice of the experimenter is, in this case, a background condition of little importance.

[30]See section 4.4.4 below for an account of Boyle's raised tube experiment.

There are, in contrast, background conditions that are of importance to the results obtained. The ambient air pressure at the time the experiment was performed, for example, is one such background condition. If the ambient air pressure had been different from the actual $29\frac{1}{8}$ inches that Boyle reported, we expect that the height of mercury required to balance out a given volume of trapped air would have been different from his recorded results. And although Boyle's table in figure 4.5 (see column C) seems to imply that the ambient air pressure remained constant during the time period over which the experiment was performed, we would not expect the same results had it been fluctuating. Even though we might expect the inverse proportionality of pressure and volume to continue to hold, we would not expect the exact same measurements as ambient pressure varies.[31]

Another background condition of particular relevance to the relationship between pressure and volume is the factor that we now call temperature. In Boyle's time, there were ways of measuring what we now call temperature, but there was not yet a standard instrument or scale. Still, Boyle did have some evidence that pressure and volume could change with degree of warmth. After describing the details of the J-tube experiment, he recounts the following trial that he performed:

> When the Air was so compress'd, as to be crouded into less than a quarter of the space it possess'd before, we tryed whether the cold of a Linen Cloth dipp'd in water would then condense it. And it sometimes seemed a little to shrink, but not so manifestly as that we dare build any thing upon it. We then tryed likewise whether heat would notwithstanding so forcible a compressure dilate it, and approaching the flame of a Candle to that part where the Air was pent up, the heat had a more sensible operation then the cold had before; so that we scarce doubted but that the expansion of the Air would notwithstanding the weight that opprest it have been made conspicuous, if the fear of unseasonably breaking the Glass had not kept us from increasing the heat (Boyle 1662, 61).

So Boyle knew that he couldn't infer beyond his actual results to a theory that those same results would hold at greater or lesser degrees of warmth. It would be an error to understand his theory as the claim that the pressure and volume of a certain body of air will always be inversely proportional, no matter what background conditions hold. In general, abstracting

[31]Boyle was indeed aware of pressure fluctuations with weather, as were all experimenters using the Torricellian apparatus; anyone who left the apparatus sitting out could easily see changes in the height of the mercury over the course of hours or days. Boyle would have known to be wary of his measurements changing with ambient air pressure, so we are probably relatively safe in assuming that he monitored air pressure to make sure that it didn't change much over the course of his experiments.

from experimental results to physical theory will involve differentiating between those background conditions under which the experimental data will be robust and those conditions under which the experimental data would not obtain. General criteria for identifying which conditions fall under which category are hard to come by, and are not the issue on which I wish to concentrate here; after all, new experiments that test variations in these background conditions can help to answer such questions. Instead, the important point is that physical theorizing—in order for it to be something *beyond* just reporting the results of actual experiments—necessarily involves making claims about the extended background conditions under which actual experimental results would still (counterfactually) hold.

4.4.3.3 Interventions. A third kind of ampliative inference is an inference beyond the interventions actually performed in the experimental procedure to different kinds of interventions (*i.e.*, interventions on different variables). We might theorize, for example, that a certain relationship or pattern discovered in a particular experiment will continue to be robust even when we intervene differently, or we might, in contrast, theorize that the relationship will break down under a different intervention.

For example, in the J-tube experiment, Boyle manipulated the external pressure on a trapped quantity of air by pouring more mercury into the long leg of the tube. The experiment constituted an intervention on external pressure and an observation of the resulting volume of air. In considering the results of the J-tube experiment and theorizing on their basis, one might speculate about whether or not the observed inverse proportionality between pressure and volume would continue to hold even if a different intervention or procedure were performed. For example, we might theorize that the relationship would continue to hold if we were to set up an experiment that intervened directly on the volume of a trapped body of air and observed the corresponding pressure that the body exerts on its surroundings. And we might, as a result of such theorizing, devise a novel experiment to test whether or not this is the case. This is, in fact, what Boyle did; he devised a new experiment that I refer to as the "raised tube experiment", which I discuss below in section 4.4.4.

Physical theorizing involves not only going beyond actual experimental results to further variable values or to extended background conditions; it also involves abstracting beyond the particular circumstances and procedures of known experiments. In fact, this is how ideas are generated for novel experiments. Speculating or making claims about how physical relationships will hold up or break down under new interventions and procedures is an important aspect of physical theorizing.

4.4.4 Boyle's raised tube experiment

Recall that the instances that Boyle had collected in the J-tube experiment were all instances of atmospheric air that had been compressed. He did not yet have any evidence about whether or not the volume and pressure of air might continue to follow the mathematical relation when it was *rarefied*.[32] Thus, from his J-tube experiment alone, he did not have evidence to confirm that the inverse proportionality of pressure and volume would hold up when extrapolated to air pressures below $29\frac{1}{8}$ inches of mercury. In addition, his J-tube experiment provided evidence only for one set of background conditions and one type of intervention (*i.e.*, an intervention on external pressure).

Boyle may have suspected that the relation he discovered could be extended to lower pressures and different background conditions, and that it would remain robust under other interventions. Physical theorizing necessarily involves making such abstractions that go beyond the experimental evidence. But novel experiments are needed to test theories that make claims about uncharted experimental territory. And this is exactly what Boyle did: he devised a second experiment different from the first. In recounting the details of the experiment, he explicitly states that he intended to test the relationship between pressure and volume when air is rarefied; as we will see, the experiment that he designed also effectively tested the relation under a different intervention.

> We made use of a Glass-Tube of about six foot long, for that being Hermetically sealed at one end, serv'd our turn as well as if we could have made the Experiment in a *Tub* or *Pond* of seventy Inches deep.
>
> *Secondly*, We also provided a slender Glass-Pipe of about the bigness of a Swans Quill [sic], and open at both ends: all along which was pasted a narrow list of Paper divided into Inches and half quarters.

[32] There was no *necessary* reason for Boyle to expect that the properties of compressed atmospheric air would obey the same mathematical relations as would atmospheric air when rarified. This point is emphasized by Agassi (2008, IV.4, "Who Discovered Boyle's Law?").

A Table of the Rarefaction of the Air.

A	*B*	*C*	*D*	*E*
1	$00\frac{0}{0}$	Subtracted from $29\frac{3}{4}$ leaves	$29\frac{3}{4}$	$29\frac{3}{4}$
$1\frac{1}{2}$	$10\frac{5}{8}$		$19\frac{1}{8}$	$19\frac{5}{6}$
2	$15\frac{3}{8}$		$14\frac{3}{8}$	$14\frac{7}{8}$
3	$20\frac{2}{8}$		$9\frac{4}{8}$	$9\frac{11}{12}$
4	$22\frac{5}{8}$		$7\frac{1}{8}$	$7\frac{7}{16}$
5	$24\frac{1}{8}$		$5\frac{5}{8}$	$5\frac{19}{20}$
6	$24\frac{7}{8}$		$4\frac{7}{8}$	$4\frac{23}{24}$
7	$25\frac{4}{8}$		$4\frac{2}{8}$	$4\frac{1}{4}$
8	$26\frac{0}{8}$		$3\frac{6}{8}$	$3\frac{23}{32}$
9	$26\frac{3}{8}$		$3\frac{3}{8}$	$3\frac{11}{36}$
10	$26\frac{6}{8}$		$3\frac{0}{8}$	$2\frac{39}{40}$
12	$27\frac{1}{8}$		$2\frac{5}{8}$	$2\frac{23}{48}$
14	$27\frac{4}{8}$		$2\frac{2}{8}$	$2\frac{1}{8}$
16	$27\frac{6}{8}$		$2\frac{0}{8}$	$1\frac{55}{64}$
18	$27\frac{7}{8}$		$1\frac{7}{8}$	$1\frac{47}{72}$
20	$28+$		$1\frac{6}{8}$	$1\frac{9}{80}$
24	$28\frac{2}{8}$		$1\frac{4}{8}$	$1\frac{23}{96}$
28	$28\frac{3}{8}$		$1\frac{3}{8}$	$1\frac{1}{16}$
32	$28\frac{4}{8}$		$1\frac{2}{8}$	$0\frac{119}{128}$

A. The number of equal spaces at the top of the Tube, that contained the same parcel of Air.
B. The height of the Mercurial Cylinder, that together with the Spring of the included Air counterbalanced the pressure of the Atmosphere.
C. The pressure of the Atmosphere.
D. The Complement of *B* to *C*, exhibiting the pressure sustained by the included Air.
E. What that pressure should be according to the *Hypothesis*.

Figure 4.6: Boyle's table giving the volume of enclosed air and the height of mercury supported in the raised tube experiment (Boyle 1662, 64). Columns A and B are measured in inches. Column C shows that the pressure of the atmospheric air was measured to be 29-3/4 inches when the experiment was performed. Column D gives the calculated pressure exerted by the enclosed rarefied air to allow the column of mercury within the pipe to balance atmospheric pressure. Column E gives the prediction of Boyle's hypothesis that the pressure exerted by the enclosed rarefied air (D) is inversely proportional to its volume (A) by some expression of the form $A \times D = \beta$.

Thirdly, This slender Pipe being thrust down into the greater Tube almost fill'd with Quicksilver, the Glass helpt to make it swell to the top of the Tube, and the Quicksilver getting in at the lower orifice of the Pipe, fill'd it up till the *Mercury* included in that was near about a level with the surface of the surrounding *Mercury* in the Tube.

Fourthly, There being, as near as we could guess, little more then an Inch of the slender Pipe left above the surface of the restagnant *Mercury*, and consequently unfill'd therewith, the prominent orifice was carefully clos'd with sealing Wax melted; after which the Pipe was let alone for a while, that the Air dilated a little by the heat of the Wax, might upon refrigeration be reduc'd to its wonted density. And then we observ'd by the help of the above-mentioned list of Paper, whether we had not included somewhat more or somewhat less then an Inch of Air, and in either case we were fain to rectifie the error by a small hole made (with an heated Pin) in the Wax, and afterwards clos'd up again.

Fifthly, Having thus included a just Inch of Air, we lifted up the slender Pipe by degrees, till the Air was dilated to an Inch, an Inch and an half, two Inches, &c. and observed in Inches and Eighths, the length of the *Mercurial* Cylinder, which at each degree of the Airs expansion [sic] was impell'd above the surface of the restagnant *Mercury* in the Tube.

Sixthly, The Observations being ended, we presently made the *Torricellian* Experiment with the above-mention'd great Tube of six foot long, that we might know the height of the *Mercurial* Cylinder, for that particular day and hour; which height we found to be 29-3/4 inches.

We can see from Boyle's data table for this experiment (see figure 4.6) that the pattern of inverse proportionality between pressure and temperature continued to hold even when atmospheric air was "stretched" or "rarefied". Thus, Boyle's theory that the relationship could be extrapolated even for lower pressures was supported by the new experimental evidence:

The proportion betwixt the several pressures of the included Air undilated and expanded, especially when the Dilatation was great (for when the Air swell'd but to four times its first extent, the Mercurial Cylinder, though of near 23 Inches, differ'd not a quarter of an Inch from what it should have been according to Mathematical exactness) the proportion, I say, was sutable enough to what might be expected... (Boyle 1662, 67).

This second experiment provided positive evidence that other abstractions beyond the original experimental data were valid as well. Not only did the inverse proportionality between pressure and volume extend to lower pressures, but it also was confirmed to hold up under slightly different background conditions. Boyle doesn't tell us much about the particular circumstances that may have differed between the two experiments, such as the clothing he may have been wearing or the time of day, but we do know that the ambient air pressure was slightly different ($29\frac{3}{4}$ inches of mercury for the raised tube experiment vs. $29\frac{1}{8}$ inches for the J-tube experiment). We might presume that there was a slight difference in ambient temperature as well between the two experiments. Although such differences in background conditions may seem insignificant, the fact that the same relation was discovered under slightly different circumstances gives us reason to believe that the original result was not a mere accident or the artifact of a unique and irreproducible set of circumstances.

Perhaps even more importantly, the second experiment involved a different intervention than the first. In the J-tube experiment, Boyle poured mercury into the long leg of the tube, effectively intervening on the external pressure bearing down on the trapped air. In response to this intervention, he observed a change in the volume of that trapped air. In the raised tube experiment, in contrast, Boyle elevated the thin tube within which he had trapped a body of air. In doing so, he was effectively intervening to increase the volume of the trapped air. As a result of that intervention, he saw the mercury level in the thin tube rise. The elevated mercury level in the thin tube was evidence that the pressure that the trapped air exerted on the mercury had decreased. Although the J-tube experiment had only provided evidence of the inverse proportionality of pressure and volume under an intervention on the external pressure exerted on a body of air, the raised tube experiment provided evidence that the inverse proportionality still held for a different kind of intervention as well—an intervention on the volume of a body of air likewise resulted in an inversely proportional change of internal and external pressure.

4.5 PHYSICAL THEORY: MORE THAN AN EQUATION

In many areas of science, and in physical science in particular, we use mathematical equations or formulas in our expression of theories. Boyle's relation of inverse proportionality, for example, can be expressed by an equation:

$$P_1 \times V_1 = P_2 \times V_2 \tag{4.2}$$

Because of the convenience and representational power of mathematical formulas as they are used in the physical sciences, it can be tempting to think that physical theories are the mathematical formulas themselves.[33] But to think this would be a mistake.

[33]Such a view was famously expressed by Heinrich Hertz in 1893:

> To the question, "What is Maxwell's theory?" I know of no shorter or more definitive answer than the following:—Maxwell's theory is Maxwell's system of equations. Every theory which leads to the same system of equations, and therefore comprises the same possible phenomena, I would consider as being a form or special case of Maxwell's theory; every theory which leads to different equations, and therefore to different possible phenomena, is a different theory (Hertz [1893] 1962, 21).

Consider for a moment what it would entail if it were true that physical theories simply *were* mathematical formulas. Obviously, a mathematical equation or formula makes an assertion about a relationship that holds among a set of variables. Assuming that one or more of those variables are continuous (as most variables in physics are), a formula summarizes an uncountably infinite set of solutions. Now, assuming appropriate physical interpretations of all of the variables and constants, what would be the most straightforward way of reading a mathematical formula as a physical theory? It would be to think of the formula as making the claim that *all and only* those combinations of values satisfying the formula represent systems that are physically possible.

But if we take physical theories to simply be mathematical formulas read in these "straightforward" terms, we will end up with most of our physical theories—especially those theories that are recognized as phenomenological—being false.[34] The "straightforward" reading of a mathematical formula as a physical theory—*i.e.*, the claim that all and only the solutions of the formula represent physically possible systems—is a *maximally permissive* abstraction from experimental data. It is the claim that the relationship specified in the mathematical formula will hold up not only in the actually tested experimental circumstances, but for *all* extrapolated and interpolated values of each variable, *all* background conditions, and *all* ways of intervening. In most cases, we know that such an abstraction is too far-reaching.

For almost any given mathematical formula that figures in one of our physical theories, we know that some sets of values that satisfy the mathematical formula do *not* represent states that are physically possible. Furthermore, we know that there are physically possible states which are *not* represented by valid solutions to the formula. For example, consider the Ideal Gas Law:

$$PV = nRT \tag{4.3}$$

were P is pressure, V is volume, n is number of moles, R is the universal gas constant, and T is temperature. The ideal gas law is an equation of state; we might casually say that

[34] This observation is related to Cartwright's criticism of the "facticity view" of laws—*i.e.*, the view that the laws of physics describe true facts about reality. See Cartwright (1983, Essay 3, "Do the Laws of Physics State the Facts?").

formula 4.3 is an empirical law that is experimentally confirmed, and that we believe it to make an approximately true claim about the set of equilibrium states that are possible for a system in the gaseous state. But what exactly do we mean when we say this? Do we really mean to make the straightforward claim that the formula itself is the theory—*i.e.*, that all and only those sets of values that constitute solutions to formula 4.3 represent physically possible thermodynamic systems at equilibrium?

Quite trivially, this is not what we mean when we call the ideal gas law "approximately true". Even considering only systems in equilibrium, there are some systems that *are* physically possible which violate the formula (*e.g.*, systems near the critical point, systems in phase equilibrium). Moreover, there will be combinations of values satisfying the formula which are *not* physically possible. For example, consider a mole of hydrogen at extremely low temperature (*e.g.*, 20 K) and four times atmospheric pressure (4 atm). Although the set of values $P = 4\ \text{atm}, V = .4103\ \text{L}, n = 1\ \text{mol}, T = 20\ \text{K}$ satisfies formula 4.3, the set of values fails to represent a real physical possibility for hydrogen (and most other gases as well), since under such temperatures and pressures hydrogen is no longer a gas (much less an ideal gas).[35] In general, there are solutions to the ideal gas law that represent systems that are physically impossible, because there will have been a phase transition or a chemical change before reaching the state matching certain temperatures and pressures.[36]

Because of the discrepancies between the set of solutions to the ideal gas law and the experimental evidence that we have for what is physically possible, it does not make sense to consider the ideal gas equation—in itself—a physical theory. Nor is the ideal gas equation unique in this way; the same will be true of almost any equation we choose to imagine as a physical theory. Therefore, if we want to preserve the judgement that any law-like equation is an approximately correct description of possible physical systems, instead of simply dismissing the equation as false because of its counterexamples, we will not want to hold the position that a formula *is* (on any straightforward reading) a physical theory.

[35]Hydrogen is in a liquid state at $P = 4$ atm and $T = 20$ K.

[36]"Exceptions" to laws of nature when read as descriptions of fact are a well-known problem that has been extensively discussed in the literature on *ceteris paribus* laws. However, there is disagreement about whether *all* laws (even those of fundamental physics) have exceptions and thus require *ceteris paribus* clauses (see Cartwright (1983, ch. 3), Lange (1993, 2000, 2002b), Pietroski and Rey (1995)) or whether fundamental physical laws are intended as exceptionless and require conditions of application rather than *ceteris paribus* clauses (see Earman and Roberts (1999)).

After all, we do not typically say, on account of "exceptions", that a mathematical formula (*e.g.*, the ideal gas law) is false, or that it should be removed from the mathematical armory of physical theory. Instead, in our hopes to arrive at a physical theory that provides an approximately true characterization of physical possibility, we craft a theory that *places limitations* on the applicability of the formulas of which it makes use.

Take Boyle's theorizing as an example. Boyle suspected that the inverse proportionality between pressure and volume might not hold up at extreme pressures. In addition, he had reason to believe that heat was likely a background factor that would cause the relation to break down. And before performing the raised tube experiment, he had no evidence that the relation would be robust when intervening on the volume of a body of air rather than on the external pressure on that body. Reasonable physical theories do not attempt to abstract universally beyond particular experimental circumstances; they are rather more modest. Further experimentation is designed to test the extent of valid abstraction.

Thus, a physical theory is more than a mathematical formula. To focus too heavily just on an abstract formula is to misrepresent the content of physical theory, because a formula or set of formulas almost always represents possibilities that we know to be false. When a theory makes use of a mathematical formula, as physical theories often do, we should understand the formula to be subject to limitations that restrict the space of physical possibilities. Sometimes these restrictions are explicit, but more often they are unstated.

The "hidden content" of physical theory is the set of things we know to be true about the set of physical possibilities that aren't representable by equations or formulas alone. One category of this hidden content is the knowledge that we have about how physical systems respond to various types of interventions. This knowledge is the *causal* content of physical theory. For example, Boyle's J-tube experiment provided evidence that, when an intervention is performed that changes the external pressure on an enclosed body of air, the volume of that body of air will change according to equation 4.2. His raised tube experiment provided evidence that, when an intervention is performed that changes the volume of an enclosed body of air, the pressure that the body of air exerts outward changes, also according to 4.2. Thus, Boyle's physical theory based on his experimental evidence included hidden causal content: he knew that the external pressure on a body of air was an interventionist cause of its volume, and that the volume of a body of air was, in turn, an interventionist cause of the internal pressure that the air exerts on its surroundings.

4.6 CONCLUSION

In this chapter, I have highlighted several interventionist causal questions and debates surrounding experiments in the early history of thermodynamics. I have also explored the process of abstraction by which experiments are represented mathematically and the ampliative inferences from those mathematical representations to physical theory. I gave an account of how physical theories can contain "hidden" interventionist causal content; generally speaking, the "hidden" content of physical theory is the set of things we know to be true about physical possibilities that aren't representable by equations or formulas alone. So even though causal relationships cannot be directly read off of the mathematical formulas utilized by our physical theories, our theories contain knowledge about how physical systems respond to various types of interventions. And this knowledge is epistemically grounded in the interventions that we have actually performed and the results we have observed in the experiments that stand as evidence for those theories.

Interventionist causal claims can also be seen in the way that we *use* and *apply* mathematical formulas when we engage in physical theorizing. In the next chapter, I will show in the context of modern thermodynamics how proper theorizing about thermodynamic systems and their environments has an inherently interventionist structure.

5.0 INTERVENTIONIST CAUSATION IN PHYSICAL THEORY

> *In most cases the scientist carves a piece from the universe and proclaims that piece* in*—namely, the* focus *of investigation. The rest of the universe is then considered* out *or* background *and is summarized buy what we call* boundary conditions*. This choice of* ins *and* outs *creates asymmetry in the way we look at things, and it is this asymmetry that permits us to talk about "outside intervention" and hence about causality and cause-effect directionality.*
>
> Pearl (2009, 420)

The previous chapters have used the interventionist account of causation to explore the connection between causal claims and physical experiments. Last chapter, I discussed the fact that the interventionist causal knowledge that is gained through experiment is not explicitly represented in the mathematical formulas of physical theory. We might wonder how, despite the "hiddenness" of interventionist causal content, interventionist structures continue to be operative in theoretical reasoning. What can we say about the meaningfulness or applicability of interventionist causal claims in physical theory?

In this chapter, I will answer the above questions in the context of modern thermodynamics. As I will show, all theorizing in thermodynamics requires careful definition of the "system" under consideration, which necessarily involves attending to the boundaries that enclose the system and the conditions imposed on those boundaries. Once boundaries are adequately specified, we end up with a strong distinction between the *internal* properties and processes of the system and those *external* influences that constrain the internal dynamics. It is in the distinction between internal properties and external influences that the natural fit between the structure of thermodynamic theorizing and the interventionist account of causation becomes apparent.

I will argue that we simply cannot theorize about thermodynamic processes without taking into account the external conditions imposed on a system. We theorize about thermodynamic processes in reference to their distance from the ideal limit of reversible processes. But since reversible processes are entirely made up of equilibrium states which, by definition, are entirely static, the notion of a reversible "process" only makes sense in reference to tiny interventions that drive a system from one equilibrium state to another. Furthermore, I will argue that it is not even possible to think about equilibrium states themselves without considering the constraints that are externally imposed on a system. I will also show that these external conditions constitute interventions which can be varied across experimental instances.

In section 5.1, I follow Rudolf Clausius' discovery and characterization of the two fundamental state functions of thermodynamics (*i.e.*, internal energy U and entropy S) by carefully examining some of his published memoirs in the period from 1850–1865. I will show that manipulated equilibrium states and carefully constrained processes are at the center of his theorizing. In section 5.2, I go on to show that interventionist reasoning continues to form the structural foundation of thermodynamic theory, and that such a foundation provides a rich basis for meaningful causal claims in the context of thermodynamic theory. As I will show, theorizing about interventions is not only a feature of the process of discovery of important elements in thermodynamic theory, but remains a structural feature of the theory itself. And since the theoretical structure of classical thermodynamics is inherently interventionist, causal claims fall naturally out of that structure. I will show that thermodynamic theory can provide clear answers to meaningful questions about whether or not a certain variable causes changes in another variable in a given context. I will also argue that, in light of the interventionist structure of thermodynamics, the consensus view of physical causation (CVPC) is inadequate for formulating or answering prominent causal questions in the context of thermodynamics.[1]

[1] For a description of CVPC, refer back to chapter 1.

5.1 CLAUSIUS' MEMOIRS, 1850–1865

In Carnot's *Reflections on the Motive Power of Heat* ([1824] 1897), he had shown that a transmission of heat from a warm body to a cooler one has the capacity to produce a certain quantity of "motive power" (*la puissance motrice*). When his work was discovered and developed by Clapeyron (1834), it came to the attention of several scientists who wished to develop the "science of heat" further (*e.g.*, Joule, Holtzmann, Thomson, Clausius). Although Carnot's theoretical results were regarded as the foundation of the science of heat by those working on developing the science, there were a number of open questions about how his principles were to be understood and how far they extended. First, there were puzzles about the relation between work and heat transfer. For example, there are obvious cases in which *no* work is produced in the process of heat transfer—*i.e.*, any case of simple heat conduction. But what happens in such cases? Where does the potential mechanical work vanish to? Another open question was how the "passing" of heat occurs. Is there an ever-constant quantity of heat whose distribution over various bodies is simply changed, or is heat actually produced and consumed? Many theorists were uncomfortable with the idea of admitting the production or consumption of heat. It seemed more theoretically elegant to presume a principle of conservation of heat, much like the principle of conservation of matter. However, some of Joule's experiments with friction and with loss of heat in cyclic processes suggested that heat actually could be produced or consumed. It was in the midst of these questions that Clausius began his theoretical work.

5.1.1 1850 paper

Clausius began his 1850 paper *On the Moving Force of Heat* by discussing Carnot's results and some of the open questions surrounding heat transfer and work. To Clausius it seemed necessary to reject the conservation of heat. Specifically, he set for himself the following task:

> These circumstances, of which Carnot was also well aware,[2] and the importance of which

[2] The "circumstances" that Clausius mentioned here refer to the open questions that I mentioned in the previous paragraph.

> he expressly admitted, pressingly demand a comparison between heat and work, to be undertaken with reference to the divergent assumption that the production of work is not only due to an alteration in the *distribution* of heat, but to an actual *consumption* thereof and inversely, that by the consumption of work heat may be *produced*.

A rejection of conservation of heat, as Clausius proposed, would require some modification of the conceptual tools that were in use. Clausius specifically identified the commonly-used concept of "total heat" as problematic when conjoined with Joule's assumption which posited the "equivalence" of heat and work. To demonstrate the problem, he stated the two assumptions as follows:

> (1) Equivalence of Heat and Work. *In all cases where work is produced by heat, a quantity of heat proportional to the work done is expended; and inversely, by the expenditure of a like quantity of work, the same amount of heat may be produced* (Clausius [1850] 1851, 4).

This assumption states, in other words, that the same proportional relation $Q/W = A$ holds, whether work is produced by a quantity of heat being expended, or whether that same quantity of heat is produced by an expenditure of work. Notice that what specifically is occurring in the "expenditure" and "production" of heat is left ambiguous in the statement; this is precisely one of the points that Clausius wished to make clear in the reasoning that will follow.

The second assumption regards the "total heat" of bodies:

> (2) Total-Heat Assumption. *"Total heat" is a term used to express the sum of the sensible and latent heat of a body and depends solely upon the present condition of that body; if all other physical properties thereof, its temperature, density, &c. are known, the total quantity of heat which the body contains may also be accurately determined* (Clausius [1850] 1851, 4, paraphrased).

The above amounts to the assumption that total heat is a state variable—*i.e.*, that a body has a property called "total heat", and that this property is entirely determined by its present state. It is a natural assumption to make if the conservation of heat is accepted, but as Clausius was about to demonstrate, it had to be rejected in order to retain (1) above while preserving theoretical consistency.

Clausius began with a simple way of seeing the tension between (1) and (2): imagine a body of gas with an initial temperature t_0 and volume v_0 that is brought to a new state with the higher temperature t_1 and greater volume v_1, then brought back to its original state. According to (2), the body has the same total heat (call it Q_0)[3] at the beginning and the end of the process, regardless of the method by which the intervening stages were achieved. Furthermore, also by (2), the intermediate state in which the gas is at t_1 and v_1 has a total heat Q_1 which is determined by some (unknown) function evaluated at that temperature and volume. Thus, regardless of the method by which the intermediate state $\langle t_1, v_1 \rangle$ is reached (call it method A), or the method by which the gas is returned from the intermediate state to the original state $\langle t_0, v_0 \rangle$ (call it method B), the quantity of heat transferred in each method will be equal in magnitude: $Q_A = Q_1 - Q_0 = -Q_B$. Now let's say that in method A, we first heat the gas to temperature t_1 at constant volume v_0 and then expand it at constant temperature t_1 to v_1. In method B, we first cool the gas at constant volume v_1 to the lower temperature t_0, then condense the gas to the volume v_0 at temperature t_0. Now Clausius noted that the work produced in method A differs from the work expended in method B, so $W_A \neq -W_B$. (Although he would prove it later in the paper, Clausius hinted here that the work produced/expended will depend on the temperature at which an expansion/compression occurs.) However, by (1), $Q_A/W_A = Q_B/W_B$, but this cannot be true if $q_A = -q_B$ and $W_A \neq -W_B$. So (1) and (2) cannot both be held true.

Therefore, to the extent that we trust (1)—and Clausius did have confidence that assumption (1) was to be preserved as the essential core of Carnot's results—we must reject the notion of total heat as a variable of state. Clausius also suggested that we must reevaluate the concept of "latent" heat, which turned out to be a misnomer on his analysis.[4] In fact, "latent" heat was supposedly heat in a hidden and insensible form, not corresponding to any temperature change of the body in question; it was a construct that functioned to save the concept of total heat. But Clausius cleared up this confusion: of the heat imparted to a system, some of it is sensible, as it may produce a temperature change in the system, but the rest of the heat imparted does not "go latent"; rather, it is converted into work and is

[3]Anyone familiar with the current use of the variable Q specifically for a heat *transfer* may find their sensibilities offended by this usage, but this is precisely the point that Clausius is working toward. Only by attempting (and failing) to work with heat as a state variable can we arrive at the conclusion that the concept of heat makes sense only in the context of a transfer.

[4]Cardwell (1971, 245) laments the fact that the concept of latent heat is still with us, despite Clausius' deconstruction of the concept.

no longer present in the system at all (at least not as heat).

In reflecting on the conversion of heat into work, Clausius distinguishes between "interior work", which he defines as "the quantity of work necessary to overcome the mutual attraction of the particles", and the "exterior work", which he defines as the quantity of work that "force[s] back an outer pressure" (Clausius [1850] 1851, 6). The former, being a function only of the attraction of the particles, is unaffected by conditions such as temperature and is invariable. The latter, however, is sensitive to conditions of temperature and pressure, and thus sensitive to the manner in which changes are made to the body in question.

Clausius then begins a mathematical treatment of these two distinct quantities of work and of heat, the last of which poses a substantial challenge because of the negative result at which he earlier arrived—*i.e.*, heat not being a state function. He begins with the Mariotte/Gay-Lussac Law: $pv = R(a+t)$.[5] Interior work, being an unknown quantity, would pose a problem if not for Carnot's idea of working with a complete thermodynamic cycle so that any interior work produced in one part of the cycle will be negated by that same quantity being absorbed in the rest of the cycle.[6]

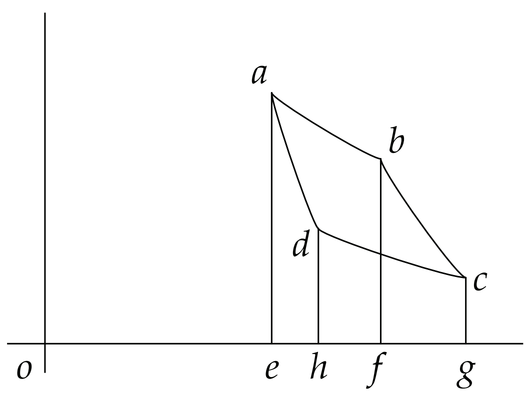


Figure 5.1: The 4-stage thermodynamic cycle Clausius considers in his 1850 paper.

[5]The modern reader will recognize this as a slightly different form of the Ideal Gas Law, where R is a constant specific to the particular type and mass of gas in question, and where temperature is expressed in degrees Celsius, with $a = 273$ as an adjustment factor.

[6]Since interior work operates, in Clausius' terms, against the "mutual attraction of the particles", and is independent of pressure and temperature along the particular path traversed, the interior work performed in a complete thermodynamic cycle that returns a body to its original state will always amount to a net zero.

Clausius chooses the cycle in figure 5.1 for consideration. The horizontal axis represents volume and the vertical pressure, with starting point a having volume oe and pressure ea. The paths on the diagram represent the following processes:

$a \to b$: Isothermal expansion carried out at temperature t.

$b \to c$: Adiabatic expansion in which the temperature falls from t to τ.

$c \to d$: Isothermal compression carried out at temperature τ.

$d \to a$: Adiabatic compression in which the temperature rises from τ to t.

Here we begin to clearly see the interventionist underpinnings of Clausius' theorizing. Note that the four points (a, b, c, d) each represent a *manipulated* equilibrium state. States such as these, as well as the processes described above, are not states and processes that arise spontaneously; they require the body of gas to be isolated from certain external influences but exposed to certain others in each of the four stages of the process described above. If a certain body of gas were at some point in time to be in one of the states represented by (a, b, c, d), it would not remain in that state or naturally transition along the curves of the diagram unless carefully guided thus under external intervention. For example, in order to achieve the process which brings the system from state a to b as described above, the gas must be in an expandable bag, but must also be in contact with a heat reservoir at temperature t that preserves the temperature of the system. In order to achieve the subsequent process from state b to c, the reservoir must be removed so that the gas can continue to expand as its temperature falls. Next, to allow the system to follow the path from c to d, a new heat reservoir at temperature τ must now be put in contact with the system while external pressure is applied. Continued pressure after that heat reservoir is removed will return the system to the original state a along path $d \to a$. Each of these interventions is stipulated in order to maintain the result that the net interior work of the procedure is zero.

In regard to exterior work, the gas produces work during the expansion phases and expends work during the compression phases. The work produced is represented by the area under the curves $a \to b$ and $b \to c$, while the work consumed is represented by the area under the curves $c \to d$ and $d \to a$. The total exterior work produced, therefore, is the area of the quadrilateral $abcd$. Now if we consider the four processes above to be infinitely small, labeling the change in volume represented by ef (see figure 5.1) as the differential dv and the change in temperature $t - \tau$ to be the differential dt, a simple geometrical calculation gives that the area of the quadrilateral is as follows:[7]

$$\text{Exterior work produced} = \frac{R\,dv\,dt}{v} \tag{5.1}$$

Clausius now turns to a calculation of how much heat is consumed in the cyclical process. Clausius defines three additional infinitesimal quantities corresponding to the volume changes of the other three stages of the process. Using these four volume differentials, he is able to write an expression for the total differential of heat consumed in processes $a \rightarrow b$ and $c \rightarrow d$.[8] We also know that, because no heat is transferred in processes $b \rightarrow c$ and $d \rightarrow a$, that the expressions for the total differential of heat consumed in each of these processes must equal zero.[9] So, using these facts and the arithmetic relation between the four differentials that

[7]See Clausius ([1850] 1851, 10) for details. Clausius uses the fact that, in the infinitely small case, one can take the quadrilateral to be a parallelogram and use the formula for the area of a parallelogram (base × height).

[8]For $a \rightarrow b$,

$$dQ = \left(\frac{dQ}{dv}\right) dv \tag{5.2}$$

where, as stated above, dv is the change in volume represented by ef in figure 5.1. And for $c \rightarrow d$,

$$dQ = -\left[\left(\frac{dQ}{dv}\right) d'v + \left(\frac{d^2Q}{dv^2}\right) \delta v d'v - \left(\frac{d^2Q}{dtdv}\right) dtd'v\right] \tag{5.3}$$

where δv and $d'v$ are the changes in volume represented by eh and hg in figure 5.1, respectively. Note that equation 5.3 is analogous to equation 5.2: the heat transfer in process $c \rightarrow d$ is mostly ruled the first term, in that the process occurs in the vicinity of point a, but involves the alternate change in volume $d'v$. The second term is a corrective term due to the process having an endpoint at $v + \delta v$ rather than v, and the third term is a correction due to the process occurring at $t - dt$ rather than t. Note also that the quantities of heat expressed in equations 5.2 and 5.3 lack a differential temperature component, since both are isothermal processes.

[9]Thus, for $d \rightarrow a$,

$$dQ = \left(\frac{dQ}{dv}\right) \delta v - \left(\frac{dQ}{dt}\right) dt = 0 \tag{5.4}$$

And for process $b \rightarrow c$,

$$dQ = \left[\left(\frac{dQ}{dv}\right) + \left(\frac{d^2Q}{dv^2}\right) dv\right] \delta' v - \left[\left(\frac{dQ}{dt}\right) + \left(\frac{d^2Q}{dvdt}\right) dv\right] dt = 0 \tag{5.5}$$

where $\delta' v$ is the change in volume represented by fg in figure 5.1. Again, equation 5.5 relates to equation 5.4 in a way similar to the relation of equation 5.3 to 5.2, with dominant terms and corrective terms. In contrast to the heat differentials for processes $a \rightarrow b$ and $c \rightarrow d$, however, these processes do involve a temperature change, and so the differential temperature components are nonzero.

he defined ($dv + \delta' v = \delta v + d' v$), Clausius reduces the expression for total heat expended in the cycle to the following:

$$\text{Heat expended} = \left[\left(\frac{d^2Q}{dtdv}\right) - \left(\frac{d^2Q}{dvdt}\right)\right] dvdt.^{10} \tag{5.6}$$

Now recall that, according to the assumption of the equivalence of heat and work, we expect that the production of a certain amount of work will be accompanied by a certain expenditure of heat, *i.e.*, $Q/W = A$, where "A denotes a constant which expresses the equivalent of heat for the unit of work" (Clausius [1850] 1851, 11). If we substitute expressions 5.1 and 5.6 for Q and W and simplify, we end up with the following equation:

$$\left(\frac{d^2Q}{dtdv}\right) - \left(\frac{d^2Q}{dvdt}\right) = \frac{AR}{v} \tag{5.7}$$

Equation 5.7 is significant because it shows that the mixed partial second derivatives of heat with respect to temperature and volume are *not* equal to each other, as would be the case if Q were expressible as a function of v and t where those two variables were independent of one another.[11] Clausius transforms equation 5.7 into the following form:

$$dQ = dU + AR\frac{a+t}{v}dv. \tag{5.8}$$

Equation 5.8 indicates that the heat absorbed by a gas can be divided into two components: the first component, U, is an arbitrary function of an independent v and t, and is therefore a state function. U represents, in Clausius' words, "the *sensible* heat and the heat necessary for *interior* work". The second component represents the heat that has gone into exterior work, and which is dependent on the path that the system takes in the process of producing that work.[12]

What did Clausius achieve in his 1850 memoir? Perhaps the most important result of the paper was his reformulation of the concept of heat: heat was no longer to be considered

[10] As the mathematical introduction of Clausius (1867b) explains, he follows Euler in using parentheses (*e.g.*, $\left(\frac{dz}{dx}\right)$) to indicate partial derivatives rather than using Jacobi's delta (*e.g.*, $\frac{\delta z}{\delta x}$). This choice is especially helpful in the derivation of this particular expression, given his use of four distinct differentials. I have followed his notation for the most part, except that I have consolidated second partial derivatives where possible.

[11] Clausius (1867b), which contains the 1850 memoir, has a mathematical introduction explaining this point at length. The footnote on page 26 is also helpful.

[12] Equation 5.8 is generally taken to be the first analytical statement of the First Law of thermodynamics (Cardwell 1971; Yagi 1981), and looks a bit more familiar if rearranged: $dU = dQ - AR\frac{a+t}{v}dv$.

something *had* by a body in virtue of its state, but rather one particular means of exchange *between* bodies when one or more of those bodies undergo changes of state. By carefully selecting a highly controlled theoretical scenario, he was able to mathematically track of all of the heat and work exchanges in which a certain thermodynamic body engaged. For our purposes here, it is important to note that it was necessary that Clausius' theoretical scenario be a product of controlled manipulation in order for him to account for all of the exchanges. The processes he describes are not real processes and would never occur spontaneously; they require external maintenance at each point. As I will show in section 5.2, these highly artificial scenarios will remain central to thermodynamic theorizing.

5.1.2 1854 paper

Clausius began his 1854 paper, *On a modified form of the second fundamental theorem in the mechanical theory of heat*, by summarizing the most important result of his 1850 paper: he had managed to show that Carnot's principle and the theorem of the equivalence of heat and work were compatible. Doing so had required the rejection of the concept of total heat, and the introduction of a new (and yet unnamed) state variable U.[13]

Clausius now steps back to consider the science of heat in more general terms. In the background of his 1854 paper is the awareness that the science of heat is entirely about "transformations". These transformations can be divided into two types: (1) simple *transfers* of heat from one body to another, and (2) *conversions* from work to heat or vice versa. Each of these two types of transformations can be carried out in what we might call either a *natural* or an *unnatural* direction. For example, we see many instances around us of heat

[13] At this point in time, Thomson had explicitly identified the importance of "making the mechanical energy of a fluid in different states an object of research", and he defined the total mechanical energy of a body as "the mechanical value of all the effect it would produce, in heat emitted and in resistances overcome, if it were cooled to the utmost, and allowed to contract indefinitely or to expand indefinitely according as the forces between its particles are attractive or repulsive, when the thermal motions within it are all stopped" (Thomson 1853, 474–476). Thomson had begun working with increments of this mechanical energy which he denoted as de. Although Clausius' dU played an identical role to Thomson's de, Clausius appears to have been somewhat uncertain as to its interpretation, and was reluctant to name it. Recall that in his 1850 paper, he only felt comfortable saying that dU accounted for sensible heat (*i.e.*, that which could be measured with a thermometer) and any internal work that was done as a body went through a change of state. It was in 1865 that Clausius named his function U "energy" for the first time (Clausius [1865] 1867). See also Cropper (1986).

transfer from a warm body to a cooler one; this is the spontaneous direction of heat transfer. Likewise, in regard to the second type of transformation, we see many instances around us of work generating heat—for example, the friction of movement spontaneously results in heat. However, it is possible to artificially force the unnatural version of each process to occur. This is exactly what a heat engine does, for example: it "converts" heat into work. Analogously, we can artificially force an unnatural heat transfer, making heat travel from a cold body to a warmer one. In these latter two artificially-induced scenarios, however, we use a "compensating" process to drive the unnatural one.[14] In a heat engine, the unnatural conversion from heat to work occurs only when accompanied by the natural direction of heat transfer from hot to cold. Likewise, a heat engine running in reverse accomplishes the unnatural direction of heat transfer (from cold to hot), but this is accompanied by the natural conversion from work to heat.

Clausius proposes his own six-stage reversible heat engine for consideration:

$a \to b$: Adiabatic expansion in which temperature falls from t to t_1.

$b \to c$: Isothermal expansion carried out at temperature t_1. Quantity of heat Q_1 is withdrawn from reservoir K_1.

$c \to d$: Adiabatic expansion in which temperature falls from t_1 to t_2.

$d \to e$: Isothermal compression carried out at temperature t_2. Quantity of heat Q_1 is deposited in reservoir K_2.

$e \to f$: Adiabatic compression in which temperature rises from t_2 to t.

$f \to a$: Isothermal expansion carried out at temperature t. Quantity of heat Q is withdrawn from reservoir K.

Like the four-stage cycle that Clausius considered in the 1850 paper, the six processes that make up this more complex cycle are processes that require maintenance by external intervention at all points (*e.g.*, by exchanges with heat reservoirs, application of pressure). Note also that, in terms of the types of transformations discussed above, Clausius has designed this cycle so as to neatly carry out two transformations: (1) a heat transmission

[14] This way of framing Clausius' work in his 1854 paper—*i.e.*, in terms of two different types of heat transformations, each admitting of a natural or unnatural direction—is due to Cropper (1986). This reasoning is also more clearly evident in the revised exposition of his "theorem of the equivalence of transformations" in Clausius (1862a), where he had the advantage of hindsight in laying out his reasoning.

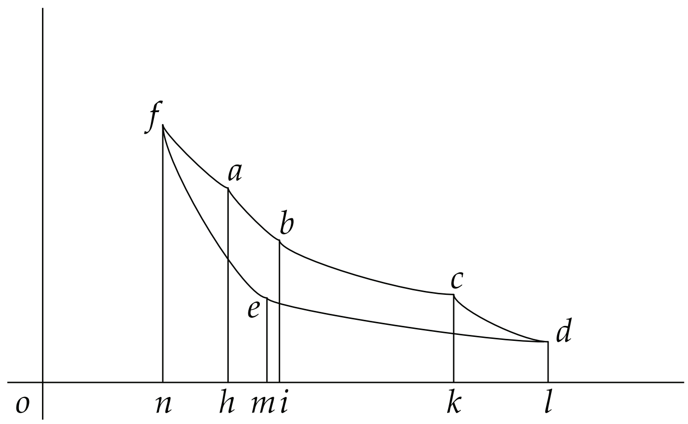


Figure 5.2: The 6-stage thermodynamic cycle Clausius considers in his 1854 paper.

in which Q_1 flows out of hot reservoir K_1 at temperature t_1 and into cold reservoir K_2 at temperature t_2; and (2) a conversion of heat Q, withdrawn from reservoir K at temperature t, into the net work W that was done by the gas over the entire cycle.

Clausius observes that in all reversible processes, there are two transformations that "compensate" for one another. He sets for himself the goal of investigating the "mutual dependence" of compensating transformations. Specifically:

> We have now to find the law according to which the transformations must be expressed as mathematical magnitudes, in order that the equivalence of two transformations may be evident from the equality of their values. The mathematical value of a transformation thus determined may be called its *equivalence-value* (Clausius [1854] 1856, 90).

Clausius' theoretical idea here is to build a metric of "equivalence values" for transformations. The values themselves are arbitrary, as long as two processes that are entirely substitutable for one another (and thus are considered "equivalent") have the same value. He starts with the following rules for his metric:

- The conversion from work to heat is positive, and the transmission of heat from a higher to lower temperature is positive. The reverse of these transformations is negative. (Thus, "natural" transformations have positive equivalence values, and "unnatural" transformations have negative equivalence values.
- The value of a transformation from work into heat must be proportional to the quantity

of heat produced (Q), and can only depend on the temperature at which the heat is exhausted (t). Thus, it must be of the following form: $Qf(t)$, where $f(t)$ is an arbitrary function of temperature.

- The value of a transmission of a quantity of heat Q from a reservoir at temperature t_1 to a reservoir at temperature t_2 must be proportional to the quantity of heat transmitted and can only depend on the two temperatures. Thus, it must be of the following form: $QF(t_1, t_2)$, where $F(t_1, t_2)$ is an arbitrary function of the two temperatures.
- In every reversible circular process, the transformations involved must compensate for one another. Thus, in a reversible circular processes involving two transformations, those two transformations must have equivalence values that are equal in magnitude but opposite in sign, summing to zero.[15]

We can now consider Clausius' six-stage process as an example in using the above rules. Recall that his six-stage process can be broken down into two transformations, each of which can be represented with an equivalence-value term. We have the equivalence-value $-Qf(t)$ for the quantity of heat Q withdrawn from reservoir at temperature t and converted into work, and we have $Q_1F(t_1, t_2)$ for the quantity of heat Q_1 transmitted from the reservoir at temperature t_1 to the reservoir at temperature t_2. And since these two processes compensate for one another in a reversible circular process, they must sum to zero:

$$-Qf(t) + Q_1F(t_1, t_2) = 0 \tag{5.9}$$

By considering another six-stage engine running in reverse and combined with his former six-stage engine, Clausius is able to arrive at the result that his arbitrary function of two temperatures $F(t_1, t_2)$ is reducible to his arbitrary function of one temperature $f(t)$ as follows:[16]

$$F(t, t') = f(t') - f(t) \tag{5.13}$$

[15]These rules are closely paraphrased from Clausius ([1854] 1856, 90).

[16]Clausius arrives at equation 5.13 in the following way. He considers a second reversible circular engine (call it engine 2) running in reverse. It transmits quantity of heat Q_1 from the reservoir at temperature t_2 to the reservoir at temperature t_1, just as the former engine (call it engine 1) would if running in reverse. However, it produces a quantity of heat Q' from work and exhausts this heat into a reservoir K' at temperature t'. For engine 2, we have the following equation:

$$Q'f(t') + Q_1F(t_2, t_1) = 0. \tag{5.10}$$

This result is important because it shows that for any reversible circular process, the equivalence values of all of the transformations can be represented in terms of conversions from work to heat or vice versa, regardless of the actual types of transformations carried out in the process. This means that the total equivalence value for any reversible circular process can take the following form:

$$\sum Q_i f(t_i) = 0, \tag{5.14}$$

Where each positive or negative quantity of heat Q_i that is exchanged with a reservoir of temperature t_i is included in the sum. And if we consider infinitesimal quantities of heat dQ, where the temperature of the reservoir to which those quantities are transferred is allowed to change, we can transform the above result into integral form:

$$\oint f(t)dQ = 0^{17} \tag{5.15}$$

And using the equivalence of mixed derivatives for a state function and equation 5.7 from his earlier paper, Clausius derived the final result that his function $f(t)$ is equal to the inverse of absolute temperature:

$$f(t) = \frac{1}{a+t} = \frac{1}{T}, \tag{5.16}$$

where $a = 273$ is the constant from Mariotte and Gay-Lussac Law. Thus,

$$\oint \frac{dQ}{T} = 0 \tag{5.17}$$

Using the fact that $F(t_2, t_1) = -F(t_1, t_2)$ and combining equations 5.9 and 5.10, we have the following result:

$$-Qf(t) + Q'f(t') = 0. \tag{5.11}$$

Then, combining the two engines by using the work produced by engine 1 as part of the required work to drive engine 2, we have the result of a net zero heat flow into each of the reservoirs K_1 and K_2, the following net transformations: a net transmission of quantity of heat Q from the reservoir at temperature t to the reservoir at temperature t', and a net conversion of work into quantity $Q' - Q$ of heat, exhausted into the reservoir at temperature t'. Therefore we have the following:

$$QF(t, t') + (Q' - Q)f(t') = 0 \tag{5.12}$$

Eliminating Q' using equation 5.11, we arrive at equation 5.13.

[17] Clausius gives $\sum \frac{Q}{T}$ and $\int \frac{dQ}{T}$ for equations 5.14 and 5.15, forms which have the misleading appearance of the contemporary form of the second law for a reversible processes. I have avoided following his usage here, because of this misleading similarity. At this point in the paper, his T represents a yet-unknown function of t that is the inverse of the arbitrary $f(t)$. It is only at the end of the paper that he is able to show that $T = a + t$ and thus $f(t) = \frac{1}{a+t}$, where $a = 273$, and thus T is absolute temperature.

One point of significance that Clausius immediately recognized about equations 5.15 and 5.17 is that, since the closed integral summed to zero, $f(t)dQ$ (or its alternate form $\frac{dQ}{T}$) had to be the differential of a state function. So in addition to the state function U that he had discovered in 1850, he had discovered a second state function, which, perhaps even more than the first, was difficult to interpret. He not only left the new state function unnamed, but also left it without a symbol.[18] The crucial point, however, remained. Clausius had realized that considerations of work and heat transfer alone are *not* enough to fully describe a process; there is another function/property that describes the "equivalence value" of a process—something like its tendency to occur in the first place.

A second point of significance is that Clausius derived equations 5.14, 5.15, and 5.17 under the condition that the processes being considered were both circular and reversible. For a circular process that is non-reversible, however, he concluded that "the algebraical sum of all transformations occurring in a circular process can only be positive" (Clausius [1854] 1856, 96). If the sum were negative, it would indicate that an "unnatural" process—either heat conduction from cold to hot, or the production of work out of heat—had occurred without a "natural" process to drive it and compensate for it. Such a state of affairs is never found to occur (whether naturally or artificially). So the new unnamed state function that he had discovered would sum to zero for any reversible cyclic process, but would sum to a number greater than zero for any irreversible cyclic process.[19]

Again we see interventionist reasoning. As in the 1850 paper, Clausius' theorizing is entirely based on consideration of a tightly controlled system and its heat/work inputs and outputs when carried through a series of changes. Manipulated systems, not natural systems, are the target *par excellence* of Clausius' thermodynamic theorizing. Notice, furthermore,

[18]See Cropper (1986) for commentary on Clausius' discovery of these two state functions and his hesitancy to assign them a name or interpretation. Clausius published a two-part paper in 1862 (see Clausius (1862a,b)) attempting to give a physical explanation of why the sum of all of the transformations occurring in a circular process should always be equal to or greater than zero. To do so, he introduced the concept of "disgregation" which referred to the dispersion of the molecules of a body. He later abandoned the concept. See Leff (1996) and Harman (1982) for commentary on Clausius' use of "disgregation".

[19]In coming to this conclusion, Clausius relies on a principle that he stated earlier in the paper: "*Heat can never pass from a colder to a warmer body without some other change, connected therewith, occurring at the same time*" (Clausius [1854] 1856, 86). The result that the algebraic sum of equivalence values must always be positive for non-reversible circular processes can be seen as the mathematical expression of the same principle.

that a very prominent feature of Clausius' theorizing up until this point has been his efforts to distinguish between state functions and path functions—*i.e.*, functions which, in contrast, are dependent on the manner in which a state is achieved. Such a distinction is inherently interventionist; it is only by theorizing about the results of various types of external interventions on a system—and carefully selecting series of interventions that were theoretically convenient—that Clausius was able to distinguish between those quantities which are independent of the manner in which a state is achieved and those which are not.[20]

5.1.3 1865 paper

Clausius states at the beginning of his 1865 paper that his goal is to provide several "convenient forms" of the fundamental thermodynamic equations which apply to specific circumstances, and much of the paper is devoted to this goal (Clausius [1865] 1867). For our purposes here, however, the most significant feature of this paper is its increased clarity over his former papers in regard to the first two laws of thermodynamics and the concepts of energy and entropy.

He first gives a restatement of the "theorem of the equivalence of heat and work", which he treated in his 1850 paper. For an infinitesimal change of condition of a body, the quantity of heat absorbed by the body will go into an infinitesimal quantity of work done on the body dW and an infinitesimal change in the state function U:

$$dQ = dU + AdW.^{21} \tag{5.18}$$

We now refer to the above as the first law of thermodynamics. Clausius then turns to the second law of thermodynamics (the subject of his 1854 paper), to which he refers as the

[20] It is only by thinking about interventions on a system that we can think about what it would mean to guide a system along *different* paths toward the same state. Two different sequences of interventions can be "equivalent" in one sense—*i.e.*, in the fact that they begin at the same initial state and end at the same final state—yet *not* be equivalent in that they used different quantities of heat or work along the way.

[21] Refer to equations 5.1 and 5.8 above. The scalar A acts as an exchange rate between units of work and "thermal units". When Clausius wishes, for the sake of convenience, to express the quantity of exterior work in thermal units, he uses the lower-case w such that $w = AW$, resulting in the expression of the first law as $dQ = dU + dw$. At times he uses the word "ergon" for w. He discusses this explicitly in the 1864 appendix to his 1862 papers in Clausius (1867b).

"theorem of the equivalence of transformations":

$$\int \frac{dQ}{T} \geq 0.^{22} \tag{5.19}$$

The two theorems above each deal with a fundamental quantity in the mechanical theory of heat. In an 1864 appendix to his 1862 papers (see Clausius (1867b, "Appendices to the Sixth Memoir (1854)")), Clausius was comfortable following Thomson in using the label "energy" for the quantity U. A portion of this energy represents the heat present in the body, which he refers to as the *thermal content*.[23] *Ergonal content* is the term that Clausius uses to account for the rest of a body's energy; by this term, he refers to the portion of the energy had by a body in virtue of the state of its particles in relation to one another.

Clausius explains that an analogous breakdown holds for the fundamental quantity involved in the second law. As I noted earlier in section 5.1.2, since $\int \frac{dQ}{T}$ is zero for any process which has the same initial and final state, Clausius reasons that $\frac{dQ}{T}$ must be the differential of a state function, and he chooses to represent this function by the variable S. S is the *transformational content* of the body, and he gives it the name *entropy*.[24] The entropy of a body, like the energy, can be analyzed into two components: the transformation-value of the thermal content and the *disgregation*, which is the transformation-value of the arrangement of particles.[25]

[22]For simplicity and consistency, I have written the integral as positive here despite the fact that Clausius writes it as negative in the 1865 paper (because he defines the heat element dQ as the quantity *absorbed* by the body). This is in contrast to his usage in 1854 and 1862, where dQ represented the quantity of heat given off.

[23]In a few of his papers, including the 1865 paper, Clausius uses the variable H to represent the quantity of heat in a body. This variable is *not* to be confused with the enthalpy function which is usually represented by H today.

[24]The name refers to the Greek word τροπὴ [transformation].

[25]Here he cites the result from his 1862 papers:

$$\int dS = \int \frac{dQ}{T} = \int \frac{dH}{T} + \int dZ, \tag{5.20}$$

where $\int \frac{dH}{T}$ represents the transformation-value of the excess of heat which the body contains in its final condition over that which it possessed in its original condition, and $\int dZ$ represents the transformation-value of the change aggregation of the particles of the body from its original to its final state. Although Clausius later abandoned his use of the concept of disgregation and it was accordingly lost as a standard technical term, Leff (1996) has argued for its pedagogical usefulness of the concept in relation to the idea of "energy spreading".

In the final portion of the 1865 paper, Clausius hints at the broader implications of the two state functions as they apply to any process whatsoever (including non-cyclical and real-world irreversible processes). As he notes, there is no change in the first law for irreversible processes; the total quantity of heat absorbed by a body will continue to equal the work done on the body over the course of the process plus the change in its energy. The only difference between an irreversible process and a reversible process is that the quantity of work for an irreversible process will be greater than that for a reversible one for the same Q.

Now what of entropy considerations in irreversible and non-cyclical processes? Recall from section that in any reversible cyclic processes, all of the transformations compensate for one another such that the total transformation-value is zero (refer back to equations 5.15 and 5.17). The change in entropy over a reversible non-cyclic process will simply be the difference between the value of S at the final state and the initial state:

$$\int_r \frac{dQ}{T} = S_f - S_0, \tag{5.21}$$

where the subscript r indicates that the integral is taken over a reversible path. The environment with which the system engages during the process will undergo a corresponding entropy change of the opposite sign. But for an irreversible non-cyclical change, there transformation will be *uncompensated* to some extent, and that uncompensated portion of the transformation will always have a positive equivalence-value over and above the equivalence-value of the corresponding reversible process:

$$\int_i \frac{dQ}{T} > \int_r \frac{dQ}{T} \tag{5.22}$$

Therefore, generally speaking, total entropy increases over the course of any irreversible process occurring in an isolated system, and this result lays the foundation for the use of extremum principles in thermodynamics, which I will discuss below in section 5.2.2. For now, I will merely hint at the connection to Clausius' reasoning above. Recall that the entropy change over the course of a process—which is the same as the equivalence-value of a process—is positive in the case of "natural" energy transformations and negative in the case of "unnatural" transformations. An unnatural transformation can only occur in conjunction with a natural transformation that compensates—in fact, overcompensates—for it. The equivalence value of a process thus represents something akin to its *tendency to occur*. A process with a positive equivalence value has a tendency to happen, and the greater the equivalence value, the greater its tendency to happen. Understood as a metric of equivalence values, Clausius' entropy thus has laid a foundation for using the entropy-maximum under a given set of experimental constraints to identify the equilibrium state toward which a process will tend. A process will continue to occur until there is no further possible change (under the given experimental constraints) having a positive equivalence value. The use of the entropy-maximization rule and other extremum principles will be discussed in more detail in section 5.2.2.

5.2 INTERVENTIONIST CAUSAL REASONING IN THERMODYNAMICS

I showed in section 5.1 that reasoning about manipulated equilibrium states and carefully-constrained processes was a central feature in Clausius' process of discovery of the central elements of thermodynamic theory. In this section, I will argue a stronger point: that manipulated equilibrium states remain of central importance in thermodynamic theory, and that these manipulated equilibrium states form a structural basis for rich interventionist causal claims.

5.2.1 The Clausius submanifold

In order to make my argument, I will be using a conceptual tool proposed by Mark Wilson that he calls the "Clausius submanifold". If we begin with a phase space—*i.e.*, a collection of all possible dynamical states of a particular type of thermodynamic system—and then consider only the set of dynamical states that are members of paths in the phase space that represent physically possible temporal evolutions, we will have only a subset of that phase space—what we might call a "primary manifold". Then, if we consider a further subset of the primary manifold—not just any state that is a member of a physically possible temporal evolution, but only those states in which the system is maintained in *manipulated equilib-*

rium—we will have what Wilson calls a "Clausius submanifold" (Wilson Forthcoming).[26]

What is a manipulated equilibrium state? Natural thermodynamic systems are in constant flux. They engage in all sorts of interactions: they transfer heat, push and pull on one another, change their volume, and chemically react. The very idea of a "system", which can only be defined by its boundaries, is already a theoretical concept that we impose on the world in order to do thermodynamic "bookkeeping" (Dill and Bromberg 2011, 93). An equilibrium state is a state of a system that is *not* undergoing a change (thermal, mechanical, or chemical), and in order for a system to achieve such a state, the system must have been allowed to relax for a sufficient amount of time without the disturbing external influences of uncontrolled contact with other systems. And such a condition requires boundaries that isolate it—or otherwise control exchanges—from other systems. Often those boundaries are put in place artificially, by human intervention/manipulation.

Consider, for example, the air in an ordinary room. If we define our thermodynamic system in relation to the walls and doors of the room, we can say that the system has a fixed volume. If no massive weather change is currently occurring, we can assume that the air pressure in the room is approximately constant (not by isolation, but by contact with an external system whose pressure is approximately constant). If some kind of air conditioning system is in place and has been running for some time, we can also say that the temperature of the room is approximately constant. We can say that most of the chemical reactions occurring in the room are in a steady state and that the concentrations of various gases are relatively uniform (except perhaps for some minor concentration gradients near any plants and/or people located in the room), with equal flow into and out of the room for each type of gas. Notice, now, that even this *almost*-equilibrium state requires artificial maintenance (the

[26]The Clausius submanifold is meant to correspond to thermodynamic configuration space (*i.e.*, a space of a much smaller dimension that is sufficient for representing any equilibrium state, with thermodynamic variables as coordinates). It is a bit tricky to define the relationship between the Clausius submanifold (which resides in the full phase space) and the thermodynamic configuration space (which is a lower-dimension space). In order to clearly define the Clausius submanifold, we will need to draw sharp boundaries around just those states in phase space which we decide "count" as equilibrium states. There are many possibilities for how to define which states will "count", some definitions allowing for fewer states in phase space that count as equilibrium states and some definitions allowing for more. We might, for example, define a state in phase space as being in equilibrium (and thus being a member of the Clausius submanifold) if its root-mean-square deviation from thermodynamic ideal energy is below a certain threshold. The choice of that threshold, however, is somewhat arbitrary.

rigidity of walls, the constant work of the air conditioner). Stricter equilibrium states require much more careful isolation and maintenance, and true equilibrium states (which only exist in theory) require idealized boundaries (*e.g.*, perfect thermal insulators, frictionless pistons, perfectly rigid containers, *etc.*).

We might also think of equilibrium as a luxury for a thermodynamic system; to be allowed such a luxury, a system must be isolated from all of the outside influences that attempt to disturb that particular equilibrium state. In practice, isolation from "disturbing" influences does not usually mean complete isolation (*i.e.*, no exchange with external systems). Rather, it requires that any exchanges with the system be controlled so that the system makes only desired exchanges, and that these exchanges be at a rate slow enough to allow the system to relax and redistribute energy uniformly. Maintenance of a desired equilibrium state is generally achieved by external intervention—by placing the system in contact with only the desired external influences and disallowing any other exchanges.[27]

We have a kind of tension, then, in the way that we think about equilibrium states. On the one hand, equilibrium states are the product of external conditions imposed on a system. On the other hand, once we consider those external conditions as given, a system will *naturally* or *spontaneously* tend toward the equilibrium state allowed by the constraints. But that spontaneous or natural behavior cannot be conceived of without external constraints being placed on the system in question. To even conceive of an equilibrium state, we must ask about the conditions imposed on its boundaries. What kind of walls enclose it? Permeable, semi-permeable, impermeable? Rigid or flexible? Adiabatic or conducting? There is no such thing as an equilibrium state unless the boundaries of the system are well-defined.[28] And those boundaries constitute external interventions on the system; they effectively *set* various thermodynamic variables to take on certain values. For example, conducting walls that put a system in contact with a thermal reservoir are effectively a way of *intervening*

[27] Wilson makes this point well:

> ... manipulated systems are generally discussed as an afterthought within typical "old mechanics" presentations ... but the same secondary status cannot be sustained within a thermodynamic setting, where heat baths, isothermal containers, diathermal partitions *et al.* comprise the very stuff of which a 'manipulation' is made (Wilson Forthcoming, 17–18)

[28] In fact, a system with no defined boundaries or external constraints is effectively a universe, and its fate is something like the "heat death" discussed by Thomson, Helmholtz, and Rankine.

on temperature. Likewise, a semi-permeable boundary is a way of selectively *intervening* on particle concentrations in the system.

Thus, equilibrium states are inherently manipulated states—manipulated to be so either by human design or by some other mechanism that effectively imposes equilibrium conditions by external intervention. Recall, from chapter 3, that an intervention directly forces a variable to take on, or remain fixed at, a certain value. Recall also that the definition of an intervention (see section 3.1.2) makes no reference to human action, and thus any entity or structure playing the role of setting certain variable values or holding them fixed can fulfill the requirements for intervention. For example, a cell membrane is a structure that effectively *intervenes* to maintain a certain equilibrium internal to the cell, by keeping interior and exterior pressures equal and by maintaining certain chemical concentrations by only allowing for select passage into and out of the cell.[29]

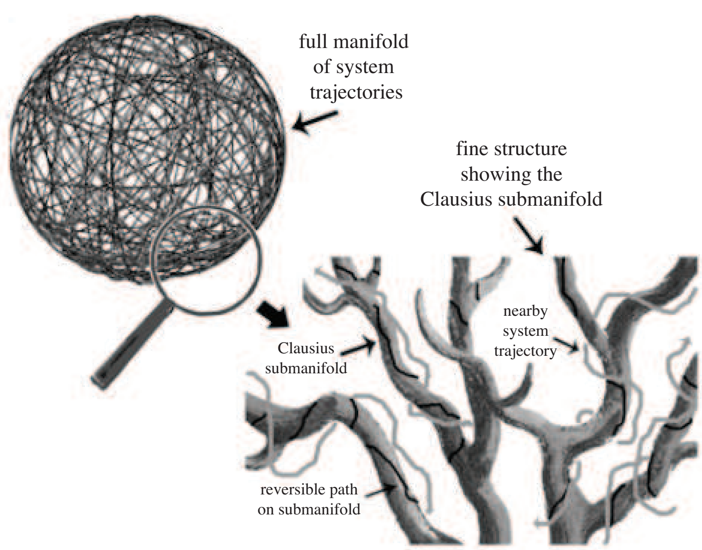


Figure 5.3: Illustration from Wilson (Forthcoming) of the relationship between the primary manifold and the Clausius submanifold. Used by permission.

Let's return to the Clausius submanifold, which is illustrated in figure 5.3. The relationship of the Clausius submanifold to the primary manifold is like that of a skeleton to flesh. The inset of figure 5.3 shows just the closed skeletal structure of the Clausius submanifold.

[29]The idea is that a cell membrane can effectively *set* the concentration of certain chemicals inside the cell by use of various mechanisms (*e.g.*, ion channels, endocytosis, *etc.*).

Many of the paths that lie entirely within the Clausius submanifold are not dynamically possible evolutions, but the detached-but-nearby gray lines are dynamically possible evolutions that are members of the primary manifold.[30] While every state on the Clausius submanifold is a manipulated equilibrium state, the paths that make up the broader primary manifold are non-equilibrium states that are less controlled and have some degree of autonomy.

But what does a Clausius submanifold do for us? Why would we want to consider only a set of states in manipulated equilibrium? As it turns out, much of our thermodynamic reasoning begins and ends with manipulated states on the Clausius submanifold. The Clausius submanifold acts as a stage of "special opportunities" from which to begin grappling with thermodynamic phenomena. The Clausius submanifold happens to be a particularly *convenient* arena for theorizing about energy and entropy, and we can use this convenient arena as a jumping-off point for theorizing about non-equilibrium states and processes. In thermodynamic theorizing we often find it especially convenient to work with scenarios that involve devices such as frictionless pistons, perfect thermal insulators, and inexhaustible heat reservoirs. Why? These imaginary devices allow us to consider certain "processes" within the Clausius submanifold under which certain thermodynamic properties remain constant. For example, a system whose enclosure features a frictionless piston can theoretically undergo an isobaric process, and considering such a system allows us to theoretically hold external pressure constant while allowing other thermodynamic properties to change. Likewise, a system put in contact with an inexhaustible heat reservoir can theoretically undergo an isothermal process, and considering such a system allows us to theoretically hold temperature constant while allowing other thermodynamic properties to change. Now, if we want to do the kind of "bookkeeping" that is important in thermodynamics—*i.e.*, if we want to investigate how various forms of energy in a system are both stored within the system and exchanged with the external environment—it makes sense to turn our attention to precisely these sorts of

[30] John Norton has called to my attention the fact that, although the physically possible evolutions depicted in figure 5.3 are shown as entirely detached from the Clausius submanifold, not all physically possible evolutions are necessarily so. There will be dynamically possible trajectories (specifically, those of systems at equilibrium) that touch or pass *through* the Clausius submanifold. In fact, if we are careful about defining the boundaries between the Clausius submanifold and the broader phase space (see footnote 26 above), there *will* be dynamically possible trajectories that remain mostly within the Clausius submanifold and occasionally (rarely) exit due to random fluctuations. Such trajectories, however, will correspond to *one* equilibrium state in thermodynamic configuration space, not to a path through thermodynamic configuration space.

processes in which all of the energy exchanges in which a system engages can be fully accounted for. These sorts of "reversible" or "quasi-static" processes are those represented in figure 5.3 by black lines traced on the surface of the Clausius submanifold.

A reversible (quasi-static) "process" is the closest thing on the Clausius submanifold to a process involving changing macroscopic states. But such a "process" really isn't a process at all: since every state that makes up such a path is an equilibrium state, no real system would move from one state on the path to the next spontaneously. In order to really think about a quasi-static process as something like a process, we must think of a system being "led"—by a series of tiny interventions—through a succession of desired states via tiny "hops".[31] In this way, we effectively "corral" a system through a sequence of equilibrium states.

Note that reversible processes were exactly what Clausius aimed to investigate when he began to consider the ways in which quantities like heat and work were exchanged and stored by a system, even if at the start of his research he was reluctant to speak generally of "energy". He designed his theoretical systems and the processes they underwent in order to fully account for all exchanges between the system under consideration and its surroundings. In fact, the reversible "processes" that constitute each stage of Clausius' four-stage and six-stage cycles (see sections 5.1.1 and 5.1.2) are lines drawn on a Clausius submanifold. These lines look something like evolutionary paths, but they do not represent real-world physical processes because the points along them actually represent static equilibrium states. Clausius was indeed fully aware of the paradoxical nature of these "processes"; he described the external forces (*e.g.*, pressure) which bring a system through such a series of changes as needing to be able, strictly speaking, to "overcome" the internal forces, yet needing to also be equal with them (Clausius 1862a, 88–89). What is needed is a series of arbitrarily tiny interventions on a system from those external forces.

We can also see, retrospectively, that what Boyle was essentially doing in his two experiments that I described in chapter 4 was identifying and recording equilibrium points on an isothermal Clausius submanifold. In his J-Tube experiment (see section 4.4.1), the thermodynamic system being studied was the body of trapped air in the short length of the tube.

[31] As Callen (1985, Ch. 4) describes, the mapping between a real sequence of states and a quasi-static path can be made to be arbitrarily close if many tiny interventions are applied to gradually changed the constraints on a system.

Since the tube was not thermally isolated, Boyle was effectively using the ambient air as a heat reservoir so that the system, when allowed to relax under the constraints he enforced, could be assumed to be at ambient temperature. The measurements that he recorded in the table in figure 4.5 thus provide pressure-volume coordinates on an isothermal Clausius submanifold. The same holds for his raised tube experiment for rarefied air (see section 4.4.4 and figure 4.6). The *real* process that Boyle carried out in his two experiments involved a series of small interventions with intermittent relaxations. Although that real process was not a reversible one, it can be thought of as an approximation to a reversible process inasmuch as Boyle "corralled" the system through this series of equilibrium states using a series of small interventions.

But the study of thermodynamics does not remain limited to the special opportunities of these strictly manipulated states. The Clausius submanifold acts as a jumping-off point for the study of *less* manipulated systems that are allowed some degree of autonomy. Paths which fall *off* of the manifold (but which are still located close to it, like the gray paths in figure 5.3) can be studied by referring back and comparing to reversible paths that are located entirely within the manifold. While states and paths that fall within the Clausius submanifold are slaves to external manipulation, paths which are located off the manifold are systems for which external manipulation is relaxed to some extent; they manifest dynamic trajectories with *some* degree of autonomy.

One way in which the Clausius submanifold proves helpful for analyzing non-equilibrium processes is that changes in state functions (like S and U discovered by Clausius) are independent of the path taken from one state to another. Thus, in order to calculate the difference in a state variable, a process on the Clausius submanifold may be used for ease of calculation, despite the fact that such a path is not actually the one being considered. In addition, the equilibrium state toward which a thermodynamic system will tend, given the experimental constraints, can be predicted using thermodynamic potential functions and experimentally-obtained state functions whose values are well-defined only in or near equilibrium (*i.e.*, on or near the Clausius submanifold).

Still, the processes that we are best able to track and describe are those which are not *entirely* autonomous—*i.e.*, those in which at least *some* variables remain constrained by manipulation. The completely autonomous evolutions of isolated systems that are familiar from physical theories in which we have complete equations of motion are somewhat foreign to thermodynamic theorizing. This is because we do *not* have a complete equation of motion for thermodynamic systems; instead, we have knowledge of certain relationships that hold for systems in equilibrium along with the laws of thermodynamics which specify certain features of thermodynamic processes.[32]

In summary, the structural foundation of thermodynamic theory is the set of equilibrium states and the reversible "processes" that can be drawn like lines through the space of such states. As I have argued here, the very idea of an equilibrium state is not possible without reference to boundaries and constraints that effectively act as external interventions that *set* the value of certain thermodynamic variables. Furthermore, we cannot think about reversible "processes", which are sequences of those equilibrium states, without thinking about a series of infinitesimal external interventions that force a system from one equilibrium state to the next. It is in this sense that interventionist reasoning constitutes the structural foundation of thermodynamic theory.

It is in this light that we can also begin to see the inadequacy of CVPC in describing thermodynamic processes. According to CVPC, causal relationships are relationships of dependency between states of a closed physical system as it evolves in time. But as I have stressed here, equilibrium states and quasi-static processes are by no means "closed" in the sense meant by CVPC.[33] Furthermore, in thermodynamics, we do not have a mathematical

[32]Gyftopoulos and Beretta (2005, 21) put this point nicely:

> In thermodynamics, the complete equation of motion is a subject of frontier research. Many attempts have been made to establish it, but the task is not yet accomplished. So the formal approach cannot be carried out. Nevertheless, many features of the equation of motion have already been discovered. These features provide not only a guidance for the discovery of the complete equation but also a powerful alternative procedure for analyses of many practical problems. The most general and well-established features are captured by the statements of the first and second laws of thermodynamics. . . . The problems that may be completely analyzed by means of the criteria and restrictions derived from the laws of thermodynamics instead of the equation of motion are those for which both the initial and final states or the *end states* of the system under study belong to a special class . . . [which] consists of all the stable equilibrium states of the system, plus all the states in which the system can be partitioned into two or more subsystems each in a stable equilibrium state.

[33]Equilibrium states require external intervention (*e.g.*, a temperature reservoir) for their maintenance, and quasi-static processes require a series of tiny interventions in addition, in order for the system to transition from state to state.

formula for stating the dependency of the state of a thermodynamic system at one point in time on the state of that same system at another point in time. Instead, we have variational principles that allow us to predict the equilibrium toward which a system will tend given the external constraints placed upon it. These variational principles allow us to describe some features of the process, but they cannot give us an exact phase-space trajectory in time. As we will see in more detail in the next section, the interventionist view of causation fits much more naturally into typical thermodynamic reasoning.

5.2.2 Thermodynamic potentials and driving forces

Let us look at some specifics for how theorizing about thermodynamic processes works. The two state functions that Clausius identified—internal energy, U, and entropy, S—constitute the central tools for theorizing about any kind of thermodynamic process. The equilibrium toward which a system will tend, given the conditions imposed on its boundaries, is governed by the energy and entropy considerations provided in the First and Second Laws of thermodynamics. The First Law tells us that any change in the internal energy of a system will be equal to the total amount of energy it gains through energy exchange with the external world, in the form of heat and/or in the form of work. The Second Law tells us that any isolated system (*i.e.*, any closed system with fixed internal energy) will tend toward its state of maximum entropy. The Second Law also has the result that the internal energy of any closed system with fixed entropy will be minimized.[34] However, neither internal energy nor entropy are directly measurable, nor do we have a specific function that tells us the dependence of U and S on other state variables. What we do have, however, are other equations of state (*e.g.*, the ideal gas law) in addition to equations for U and S in *differential* form, which tell us about the way in which small changes in other state variables relate to small changes in energy and entropy:

$$dU = TdS - pdV + \sum_j \mu_j dN_j \tag{5.23}$$

$$dS = \left(\frac{1}{T}\right) dU + \left(\frac{p}{T}\right) dV - \sum_j \left(\frac{\mu_j}{T}\right) dN_j, \tag{5.24}$$

[34] For a clear exposition of this point, see Callen (1985, Ch. 5).

where T is absolute temperature, p is pressure, V is volume, μ_j is the chemical potential for species j, and N_j is the number of particles for species j.[35]

Notice that each term in both equations above involves a pair of conjugate variables. The second term in equation 5.23, for example, involves pressure and volume as a conjugate pair. For every pair of conjugate variables, one of the variables is extensive (*i.e.*, additive such that the property of a system is equal to the sum of that property for all of its component subsystems), while the other is intensive (*i.e.*, independent of the size of the system). Furthermore, depending on the factors controlled in a given experimental context, each pair of conjugate variables tells us something about a tendency of the system as it moves toward equilibrium in that context. Since conjugate variables will be extremely important for our purposes here, let's concentrate on one pair and use an example to decipher its practical meaning.

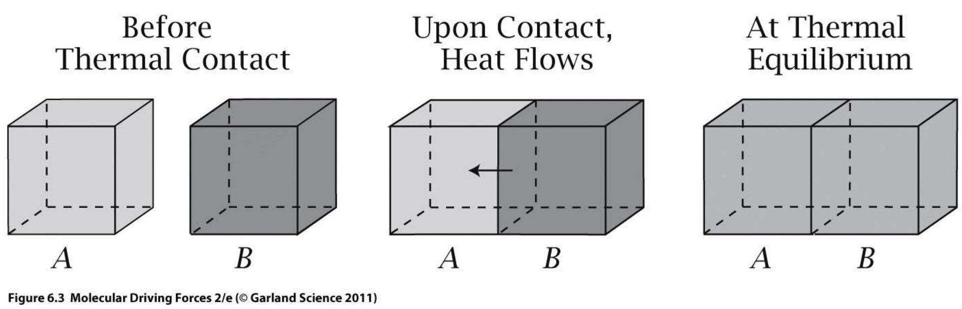


Figure 5.4: Two thermodynamic systems A and B before, during, and after arriving a thermal equilibrium process. From Dill and Bromberg (2011, 100). Reproduced with permission of Garland Science.

Let's concentrate on the term $\left(\frac{1}{T}\right) dU$ in equation 5.24. To understand this term, consider the process pictured in figure 5.4. We begin with two systems A and B, each enclosed in a rigid container. System A begins at temperature T_A and system B at T_B, where $T_A \neq T_B$. The two systems are then brought into thermal contact with each other, but remain thermally insulated from the rest of the world. Now each system has an unknown entropy that can

[35]While the first two terms of both equations may look somewhat familiar from the discussion of Clausius' theoretical work above (sections 5.1.1–5.1.3), the third terms represent chemical potentials as driving forces for particle exchange, which we owe to the work of Gibbs. Gibbs' work on chemical potentials postdates Clausius' foundational work discussed above.

be expressed as a function of its internal energy, volume, and particle numbers, and since entropy is an extensive quantity, the total entropy of the combined system can be expressed as $S_{Total} = S_A(U_A, V_A, \mathbf{N}_A) + S_B(U_B, V_B, \mathbf{N}_B)$. Since entropy will be maximized at equilibrium, we use equation 5.24 to write the differential expression for S_{Total} and set it to zero:

$$\begin{aligned} dS_{Total} = \left(\frac{1}{T_A}\right) dU_A + \left(\frac{p_A}{T_A}\right) dV_A - \sum_i \left(\frac{\mu_{Ai}}{T_A}\right) dN_{Ai} + \\ \left(\frac{1}{T_B}\right) dU_B + \left(\frac{p_B}{T_B}\right) dV_B - \sum_j \left(\frac{\mu_{Bj}}{T_B}\right) dN_{Bj} = 0 \end{aligned} \tag{5.25}$$

If we assume that there is no particle exchange between the two systems and that no chemical change occurs within each system, we can eliminate the terms that allow for changing particle numbers. And since the containers are rigid, we can eliminate the terms that allow for changing volume. Furthermore, given that the combined system is isolated from the external world, the total internal energy of the combined system must remain constant, and any change in energy of either system must be compensated by a change in energy of the other. Thus, $dU_A = -dU_B$. So we have the following simplified expression:

$$dS_{Total} = \left(\frac{1}{T_A} - \frac{1}{T_B}\right) dU_A, \tag{5.26}$$

which will be equal to zero when $T_A = T_B$.

Thus we have derived the well-known result that two objects brought into thermal contact will reach equilibrium when their temperatures are equal. But more importantly for our purposes here, we can interpret the factors in equation 5.26 in light of this equilibration process. The difference in temperatures between the two systems leads to a nonzero value of the factor $\frac{1}{T_A} - \frac{1}{T_B}$, which effectively acts as a "force" driving a change in dU_A. More generally speaking, when a system is placed in thermal contact with a system at a different temperature, the temperature difference between the two systems acts as a force driving an exchange of energy between the systems. Phrased in terms of a system and its surroundings, $\frac{1}{T}$ describes the tendency of a system to exchange heat with its environment; it is the incremental relaxation that a system experiences in transferring a small bit of its energy dU.[36]

Physicists commonly use the language of "driving forces" in referring to each of the intensive parameters in the fundamental thermodynamic equations. Looking back again at the fundamental equations (equations 5.23 and 5.24), a difference between the pressure p of the system and its environment will act as a driving force for exchanges of volume dV between the system and its environment, and a difference between the concentration of a particular species μ_j in the system and its environment will act as a driving force for exchanges of particles of the respective species with the environment (dN_j). The force or tendency represented in each of the conjugate pairs (T, p, μ) can act, separately or together (depending on the circumstances), to drive changes in its paired variable (dU, dV, or $d\mathbf{N}$, respectively), and thus to drive the system and its environment toward the equilibrium state of maximum entropy.[37]

This "driving force" language—and its basis in the way in which the environment exchanges energy and entropy with a system—is entirely corroborated by the interventionist

[36]Alternatively, we could have begun with the fundamental equation for internal energy (equation 5.23) to derive the same result. Internal energy, like entropy, is an extensive quantity, and so $U_{Total} = U_A(S_A, V_A, \mathbf{N}_A) + U_B(S_B, V_B, \mathbf{N}_B)$. The energy minimization principle that results from the Second Law tells us that the equilibrium state will be the state which minimizes the internal energy for a given total entropy. Thus, even though our composite system is closed and the total energy of the system is fixed, we can think of equilibrium as the state that *would* minimize the internal energy if energy exchange with the environment were permitted and total entropy were held fixed. For fixed total entropy, $dS_A = -dS_B$, and we end up with the following equilibrium expression:

$$dU_{Total} = (T_A - T_B)dS_A = 0 \tag{5.27}$$

Under this alternate representation, the difference in temperature $T_A - T_B$ is the driving force and dS is the quantity that is changing. The energy minimum principle is analogous to d'Alembert's principle of virtual work in mechanics: in our example here, if we imagine a tiny "virtual" change in the entropy dS_A, dU_{Total} represents the work that would be done by the force $T_A - T_B$ acting to produce that change. At equilibrium, where $T_A = T_B$, the quantity dU_{Total} will be zero because the quantity of work done by the force $T_A - T_B$ during any virtual change dS_A would be equal to zero.

Physicists use the language of "driving forces" in both the entropy and energy representations. When we flip between the energy picture of a system and the entropy picture of that same system, the metric by which we measure progress toward equilibrium changes. Each metric has its own way of characterizing the driving force because, in changing our metric of progress, there is a transformation on the force term. Still, physically, it is one and the same force driving the system toward equilibrium. This happens to be a case of a widely-noted feature of the interventionist account of causation: when we change the set of variables with which we characterize a causal system, our characterization of the causal relationship itself can change.

[37]Both Callen (1985, Ch. 2) and Dill and Bromberg (2011, Ch. 6) walk through the three cases of thermal equilibration, mechanical equilibration, and chemical equilibration. In each of these cases, we isolate a different driving force in the fundamental entropy equation.

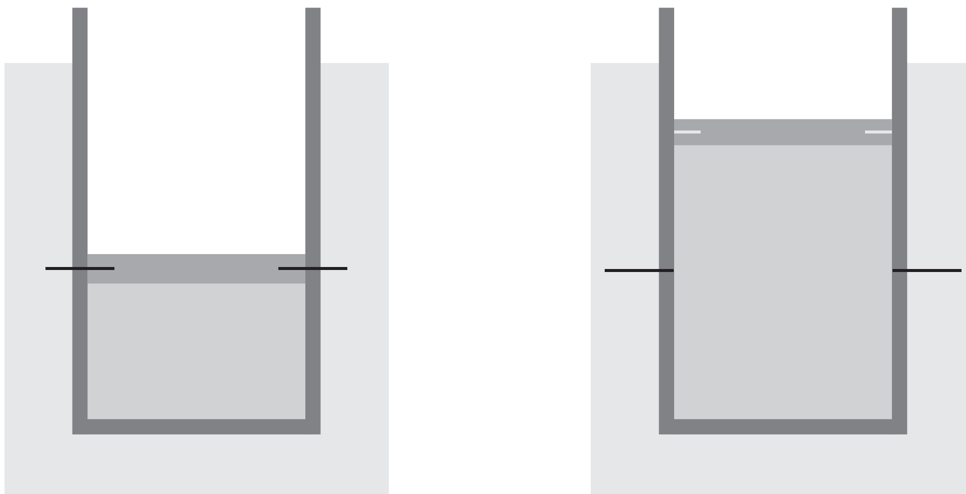
account of causation. Recall the structure of an experimental series that we defined in section 3.2: we have evidence for the claim that a variable X is a interventionist cause of another variable Y if and only if we have an experimental series in which, by intervening differently on X while keeping all other background conditions $\mathbf{BC}$ and auxiliary interventions $\mathbf{S}$ the same, we see associated variation in the value of Y. Let's now consider the temperature equilibration case above as an experimental series, with system A as the causal system under investigation.

The intervention in question was that of placing system B in thermal contact with system A. We considered this intervention specifically in conjunction with auxiliary interventions that maintained system A at constant volume V_A and constant particle numbers $\mathbf{N_A}$. The intervention led to a change in T_A, since the original temperatures of the two systems were not equal, and this change resulted in a change in the internal energy of the system (U_A). Now, if we contrast experimental instances in which we vary our intervention by putting system A in contact with a system at varying temperatures $T_{B1}, T_{B2}, \ldots T_{Bn}$, while still holding V_A and $\mathbf{N_A}$ constant at the same values, we will note that there are corresponding variations in U_A. Therefore, the interventionist account confirms that the temperature T of a system is a cause of its internal energy U. Interventions on temperature lead to changes in internal energy via exchange of heat, and this seems to be precisely what physicists mean to convey when they use "driving force" language with respect to temperature.

The intervention in the above case, where we have an equilibration process between two finite systems with differing initial temperatures, is an example of a "soft" or "parametric" intervention in that it *modifies* the temperature of our system rather than determining it completely.[38] As we saw above in Clausius' reasoning, thermodynamics also provides conceptual tools for theorizing about "hard" or "structural" interventions that entirely determine the value of an intensive parameter for a system. We call these theoretical entities "reservoirs" or "baths", and they have the property of being able to exchange one or more extensive quantities while their corresponding intensive properties remain constant. For example, an energy bath (*i.e.*, a temperature reservoir), by virtue of its size, is able to exchange energy with a system with which it is put in contact with negligible effect on its tempera-

[38] Refer back to section 3.2.2; see also Eberhardt and Scheines (2007) and Eberhardt (2007).

ture. Likewise, a volume bath (*i.e.*, a pressure reservoir) is able to exchange volume while remaining at constant pressure, and a particle bath (*i.e.*, concentration reservoir) is able to exchange particles while maintaining constant particle concentrations. When we consider theoretical experiments in which we put a system in contact with a reservoir instead of a finite system, we are effectively considering a hard intervention that *determines* the value of the relevant intensive variable in our system. Such theoretical experiments bring the interventionist causal structure into even clearer relief: setting the value of an intensive variable in a system changes the corresponding extensive variable.



(a) (b)

Figure 5.5: Illustration of a pressure-driven process, depicting (a) the system in its initial equilibrium state before the piston-locking pins are released; (b) the system once it has reached its new equilibrium state after the pins are released. This image shows the result of the case where $p_0 > p_{Res}$ and the piston rises, but all of the same considerations would apply in the case that $p_0 < p_{Res}$ and the piston falls.

Let's look at just such an example for a different pair of conjugate variables: volume and pressure. Consider a system that is in an initial equilibrium state $(p_0, T, \mathbf{N})$. Suppose that we intervene on the system by bringing it into contact with a reservoir that maintains the same temperature T as the system but a different pressure p_{Res}. We might do so by releasing an initially-locked piston, allowing it to move freely between the system and the reservoir (see figure 5.5). The process that ensues will be ruled by a maximization of the entropy of the total combined system, so we are interested in the condition where $dS_{Total} = 0$:

$$dS_{Total} = \frac{1}{T_{Res}} dU_{Sys} + \frac{p_{Sys}}{T_{Res}} dV_{Sys} + \frac{1}{T_{Res}} dU_{Res} + \frac{p_{Res}}{T_{Res}} dV_{Res} = 0 \tag{5.28}$$

Due to conservation of volume and conservation of energy, $dU_{Sys} = -dU_{Res}$ and $dV_{Sys} = -dV_{Res}$, so the above condition reduces to the following:

$$dS_{Total} = \left(\frac{p_{Sys} - p_{Res}}{T_{Res}} \right) dV_{Sys} = 0 \tag{5.29}$$

We can see here that it is the pressure difference between system and reservoir that is driving the exchange of volume. And this physical interpretation in terms of driving forces matches the interventionist causal interpretation. By placing the system in contact with the reservoir, we effectively *set* the pressure of the system to a new value, and this intervention was accompanied by a change in volume. Were we to impose a different pressure on the system by placing it in contact with a reservoir at a different pressure, we would see the corresponding volume change as well. Thus, pressure is a interventionist cause of volume.

The agreement between the interventionist account and the driving-force language also extends into the realm of irreversible thermodynamics. In irreversible thermodynamics, a system is broken down into subsystems, each of which is considered to be in local equilibrium. We can then speak of a "generalized force" or "affinity" that drives changes in each extensive parameter of each subsystem. For example, take the simplest possible non-equilibrium case, where we have a composite system composed of two subsystems, each in different equilibrium states. The generalized force $\mathscr{F}$ acting on a particular extensive variable of one of the subsystems is defined as the partial derivative of the total entropy with respect to that variable. For example, consider a composite system in which each subsystem has the same temperature and particle concentrations, but a different internal pressure. From physical intuition, we expect that the volume of the two subsystems will change as a result, but the explanation from entropy production proceeds as follows. The generalized force driving a change in the volume of one subsystem (call it subsystem A) is the partial derivative of the entropy of the total system with respect to the volume of subsystem A:

$$\mathscr{F}_V \equiv \left(\frac{\partial S_{Total}}{\partial V_A} \right)_{V_{Total}} = \frac{\partial S_A}{\partial V_A} - \frac{\partial S_B}{\partial V_B}, \tag{5.30}$$

which, referring back to equation 5.24, is simply the same thing as the difference in the intensive parameters for the two systems:

$$\mathscr{F}_V = \frac{\partial S_A}{\partial V_A} - \frac{\partial S_B}{\partial V_B} = \frac{p_A}{T_A} - \frac{p_B}{T_B}. \tag{5.31}$$

Now, this "force" acts on the "flux" of its respective extensive variable, where the flux for the volume of subsystem A would be defined as its time rate of change:

$$J_V = \frac{dV_A}{dt} \tag{5.32}$$

And if we multiply that force times the flux, we get the time rate of change of the entropy due to the volume changes within the system:

$$\frac{dS_{Total}}{dt} = \frac{\partial S_{Total}}{\partial V_A}\frac{dV_A}{dt} = \mathscr{F}_V J_V \tag{5.33}$$

On an interventionist causal analysis, the above line of theorizing can be framed entirely in terms of subsystems intervening on one another and each subsystem responding in turn to those interventions according to its internal causal structure. In fact, the most important key to successful thermodynamic theorizing is the careful definition of the boundaries between systems/subsystems and accounting for the transactions that occur at those boundaries. Interventionist reasoning fits naturally into thermodynamic theorizing because its distinction between the interventions external to a causal system and the causal relations internal to that system is perfectly applicable where thermodynamic boundaries are well-defined. Since interventions are always performed *on* a causal system from outside, it is entirely natural to label exchanges between a system and its environment as interventions of the environment on those systems.

In the previous section, we already saw that the structure of thermodynamic theory and the centrality of manipulated equilibrium states is better aligned with the interventionist account of causation than with CVPC. We can now see additional ways in which CVPC is inadequate in describing thermodynamic theorizing.

First, CVPC is not equipped to characterize a single variable as a cause of another. On CVPC, causes can only be entire thermodynamic states, since no subset of a state description allows for a prediction of the future evolution of the system. Yet, as I have shown above, appropriate application of thermodynamic principles provides us with reason to identify certain variables as driving forces of changes in certain others, and these same variables are those identified as causes under interventionist analysis. In contrast, CVPC provides us with no rationale for picking out a particular variable over any other as the cause of a change in another; for CVPC, only entire states can be causes and effects.

In addition, as we have seen, the definition of systems and their boundaries is absolutely central to thermodynamic theorizing. These boundaries provide us with the important distinction between the external influences on a system and the internal dynamics of the system (*i.e.*, how that system responds to those external influences). Yet CVPC relies entirely on an *autonomous* model of dynamics in which everything about the evolution of a system can be predicted purely by knowledge of its beginning state and its internal dynamical rules. For a theory such as thermodynamics, where we have no fully-developed autonomous dynamics, CVPC is wholly inadequate.

5.3 CONCLUSION

I have argued in this chapter that interventionist reasoning is evident not only of the process of discovery of thermodynamic theory, but in the very structure of the theory itself. We can see the interventionist underpinnings in the Clausius submanifold that forms the skeleton of thermodynamic theory and in the "driving forces" which turn out to be interventionist causes of their respective conjugate variables. Interventionist causal claims, in which one variable is said to cause another in a given context, can be formulated and assessed quite naturally using standard "textbook" thermodynamic language and explanations.

6.0 PROSPECTS FOR AN ANTI-FOUNDATIONALIST, ANTI-REDUCTIONIST UNIFICATION OF CAUSATION

> *. . . the widely accepted idea that all true causal claims in common sense and the special sciences 'instantiate' fundamental physical laws which are causal in character, with [the] causal status of the former being 'grounded' in these fundamental laws alone is deeply problematic.*
>
> Woodward (2007, 70)

In this dissertation, I have argued that the interventionist account of causation deserves attention as a candidate for a "physically respectable" account of causation—*i.e.*, it explicates important parts of the experimental practice of physics and important aspects of the ways that physical theory is used and applied to concrete problems. In this concluding chapter, I will begin by summarizing the arguments I have made so far and then discuss some prospects and implications of my thesis.

I began in chapter 1 by criticizing the consensus view of physical causation (CVPC). Although it fails in several ways that have already been identified in the philosophical and scientific literature, my novel criticisms were: (1) that it fails to provide an epistemology which connects causal knowledge in physics to the experimental methodology of physics; and (2) that it fails to give a satisfactory account of causal relationships in thermodynamic theory. I then proposed the interventionist account of causation as an alternative because it can succeed where CVPC fails.

In chapter 2, I examined the relationship between scientific experimentation and causal theorizing. I argued that the close link between experimentation and theorizing about causes is not a new idea; it dates back to the beginnings of the scientific revolution. In its most basic form, the idea is the following one: in order to investigate the cause of a particular result, a scientific researcher must intentionally set up various scenarios, all of which differ in only one respect; if the result differs in each scenario, then the cause of that difference can only be that which differed in the original setup. I examined the history of this methodological principle in the thought of some of the major "founding fathers" of experiment: Galileo Galilei, Francis Bacon, and John Stuart Mill.

In chapter 3, I gave a philosophical analysis of the link between experiment and the testing of causal claims. My analysis was rooted in the interventionist account of causation, which I see as a developed and sophisticated realization of the historical thread of thought examined in chapter 2. The basic tenet of the interventionist account is that a causal relationship is one such that, under at least some circumstances, one variable (*i.e.*, the effect) can be varied or changed by intervention on another variable (*i.e.*, the cause). Under this understanding of a causal relationship, scientific experimentation becomes a straightforward method for the empirical testing of causal claims. Still, not all experiments—particularly those in physics—are performed with the goal of testing a causal claim, so I gave a general account that identifies those features of an experiment that suffice for it to afford causal inference. I showed (1) that the set of causal inferences afforded by any experiment is determined solely on the basis of contrasting case structures that I called "experimental series" and (2) that the conditions that suffice for causal inference obtain quite commonly, even among "ordinary" scientific experiments not explicitly designed for the testing of causal claims.

In chapter 4, I examined the early history of thermodynamics as a case study in causal reasoning and the relationship between theory and experiment. I discussed the experiments of Torricelli, Roberval, Pascal, and Boyle. I showed that the major discoveries of the period—the spring of the air and Boyle's Law (describing the functional relationship between the pressure and volume of a gas)—were made against a rich background of causal debate and arguments using causal-experimental contrasts. Specifically, I showed that the experiments, and the arguments of the scientists themselves about their own experiments, made use of experimental series matching the schema for testing causal claims that I characterized in chapter 3. I also showed how, in Boyle's experiments and his discovery of the relationship between volume and pressure, the seeds of thermodynamic theory—and its hidden causal content—were already being sown.

Finally, in chapter 5, I turned to modern thermodynamic theory. I showed that manipulated equilibrium states and carefully constrained processes were at the center of Clausius' theorizing as he developed the concepts of internal energy and entropy. I then argued that such reasoning continues to be central in thermodynamics today: all theorizing in thermodynamics requires careful definition of the "system" under consideration, which necessarily involves attending to the boundaries that enclose the system and the conditions imposed on those boundaries. By defining boundaries, we create a theoretical distinction between internal properties and external influences, and that distinction provides a natural home for interventionist analysis, where we can then use external interventions to tease apart internal causal relationships.

In summary, I have argued for the applicability and suitability of the interventionist account within the domain of experimental and theoretical thermodynamics. In this concluding chapter, I would like to take the above arguments one step further and discuss the potential philosophical benefits of an interventionist understanding of physical science. In section 6.1, I consider the prospects for arriving at an interventionist understanding of causation in subfields of physics other than thermodynamics. In section 6.2, I describe what I foresee to be the biggest potential philosophical benefit of an interventionist account of causation in physics.

6.1 PROSPECTS FOR AN INTERVENTIONIST UNDERSTANDING OF CAUSATION IN OTHER SUBDOMAINS OF PHYSICS

In order to assess the prospects for an interventionist analysis of other areas of physics beyond thermodynamics, let's begin by focusing on those features of thermodynamic theorizing that make it particularly conducive to interventionist reasoning. In order to begin an interventionist analysis in any domain, we must define our "causal system". Once this is decided, everything external to the causal system fits into one of two categories: (1) those things that influence the system and thus are considered external interventions, and (2) those things which do not influence the system and thus are considered irrelevant to the internal workings of the system. The reason that thermodynamic theorizing is so conducive to interventionist reasoning is that it explicitly and purposefully proceeds in this fashion: it clearly defines the system under consideration and specifies the conditions imposed on its boundaries. The boundaries assigned to the system effectively decide the way(s) in which the system does and does not interact with its environment. In fact, the importance of system boundaries in thermodynamics is made evident by the fact that there are several specialized terms for various types of systems and boundaries (*e.g.*, open systems, closed systems, isolated systems, adiabatic boundaries, semipermeable boundaries, *etc.*).

I propose that the key to an interventionist causal analysis of other subfields of physics will lie in the following question: within the given subfield of physics, is there a way to partition the object of physical analysis into *system* and *environment*? I suspect that the answer to this question will be "yes", for most (if not all) subfields. Of course, in many subfields of physics, this act of partitioning is *not* the customary way of working with the physical theory in question; this is perhaps the most significant difference between thermodynamics and other subfields. However, the fact that it is not customary does not mean that it is not possible or beneficial to do so.[1] Let's look at an example to see how this might work.

Consider the well-known physics demonstration involving two parallel wires, each carrying current. If their current is flowing in the same direction, the two wires will attract each other; each current produces a magnetic field that is perpendicular to the current flow in the other, and the force on that wire due to the magnetic field will be toward the other wire. It is a relatively simple matter to consider the fields and forces involved, and considering both wires as a collective system, we might say that the behavior of the two wires is a result of the system simply satisfying the laws of electromagnetism and mechanics. No reference to causation is necessary. However, if we consider one wire and a certain fixed volume around it as the "system" and everything else (including the other wire) as part of the environment, we can begin to ask interventionist questions. We can ask, for example, how the behavior of the wire within our system does or does not change in response to varying external stimuli (for example, we might consider changes in the magnitude or direction of the current in the

[1] By "beneficial" here, I am referring to the potential philosophical benefits of a unified account of causation, which I discuss below in section 6.2.

second wire). Such interventions change the magnetic field within the volume of our system, and those changes in magnetic field result in different behaviors of the wire. Thus, we can say that the magnetic field is an *interventionist cause* of the trajectory of the wire, and we can even mathematically characterize the causal dependence of that trajectory on the magnetic field. It is only when we "carve up" a physical scenario into *system* and *environment* that we are able to use the interventionist account to give a clear and precise characterization of which relationships are causal and which are not.[2]

This example would need to be elaborated and additional examples would need to be provided to more adequately assess the promise of the interventionist account of causation in other areas of physics. However, the main point is this: a meaningful interventionist analysis requires that our system be isolated to a certain extent, so that it has a reasonably independent internal dynamics, but that it also be embedded within an environment in such a way that the environment is able to provide a source of external interventions (see Woodward (2007, 92), Woodward (2013, §12), Pearl (2009, Epilogue)). To the extent that we can create such a partition into system and environment within a subdomain of physics, we will be able to subject that subdomain to interventionist causal analysis. I see no obvious reasons to doubt that such a partition will be possible in most, if not all, subdomains of physics.[3]

6.2 A UNIFIED CONCEPT OF CAUSATION

At this point, we might wonder: *if* the interventionist account of causation did prove to be appropriate and applicable to most (or all) areas of physics, what would be the philosophical pay-off? In the preceding chapters, I have already argued for two benefits: (1) that the interventionist account of causation provides an epistemology of causal knowledge in physics

[2] See Pearl (2009, Epilogue), who explains the necessity of "carving up" the universe in order to see causal relationships.

[3] Woodward (2013) expresses doubts about the applicability of the interventionist account on very small and very large scales. For reasons I will explain below, I do not see scale in itself as a problem. As I see it, the question is entirely reducible to whether or not we are able to define an appropriate partition between system and environment.

that is rooted in experiment; and (2) that the interventionist account of causation is well-suited to the patterns of reasoning that are intrinsic to thermodynamic theory. The third potential philosophical benefit is the prospect for a unified concept of causation. I will now give a sketch of the kind of unification that the interventionist account may be able to offer.

I have already proposed that there is a unified concept of causation in physics that extends across both theory and experiment. In the experimental realm, a cause-effect relationship between a variable X and another variable Y exists when intervening to vary X reveals associated variations in Y when all other factors (**BC** and **S**) remain constant. In the theoretical realm, the same definition of a causal relationship is operative when we draw distinctions between a system's internal relationships and the external influences on that system. A causal relationship between two variables X and Y *internal* to the system holds when, under some set of externally-imposed constraints, an *external* influence varies the value of X and the theory predicts a resulting change in Y. Furthermore, the unity of the concept of causation across the experimental and theoretical domains is the basis for an epistemological account of causal knowledge in physics. Once we draw certain boundaries around a physical system, we know what kinds of external influences are and are not allowable, and what kinds of internal behavior we expect the system to exhibit as a result. That knowledge is causal knowledge, and we owe that causal knowledge to collective experience from experiment.

This result is important for the philosophy of physics, but it is also relevant to general philosophy of science. If my argument continues to hold for other experiments and other theoretical subfields of physics beyond thermodynamics, it has great significance in that *causation would not be something wholly different in physics than it is in other sciences*, contrary to the presumed peculiarity of "physical causation" that CVPC attempts to characterize. As shown in chapter 3, causal analysis methods in physics, especially in experimental settings, are understandable in terms similar to those of causal analysis methods in the "higher-level" and "special" sciences. And the unification potentially extends even further: on the view presented in this dissertation, we have opened doors for a possible unification of so-called "physical causation" and our everyday concept of causation.

What kind of unification? The unification offered by an interventionist understanding of causation in physics is a very different unification than might be expected by someone who subscribes to CVPC. Very often, CVPC is allied with a *foundationalist* view of causation. Causal foundationalism (a position named by Ney (2009)) is the view that physical causation is fundamental and that causal facts at higher levels (*i.e.*, causal facts as pertain to everyday life and the the "higher-level" sciences) are *dependent on* or *grounded in* facts about physical causation.[4] Causal foundationalism sees itself as providing a certain "unification" of physical causal relationships and causal relationships at higher levels; it proposes to do so by postulating some kind of dependency relation between "higher-level" causal relationships and "fundamental" physical causal relationships. On my view, however, causal foundationalism is an unsatisfying way of unifying causation under a single account. As I argued in chapter 1, there are abundant reasons to reject CVPC as a view of physical causation. Furthermore, even if it might be reasonable to ask for "truth-makers" or a "grounding" of higher-level causal relationships at some lower level, there seems to be no reason to expect that the "grounding" must be in the form of a distinct account of causation that is peculiar to physics.

If we relinquish the expectation that there should be an account of causation peculiar to physics and that such an account of "physical causation" will provide a foundation for the rest of our causal reasoning, it becomes perfectly legitimate to label and identify certain relationships in physics as causal using an independently established and experimentally-centered understanding of causal relationships. According to the arguments put forward in this dissertation, there are prospects for *this* type of causal unification, which I will call an *anti-foundationalist* unification of causation. I will now attempt to sketch a view of this sort.

[4]Despite a similar name, causal foundationalism is not to be confused with a position that John Norton calls *causal fundamentalism*, which is the view that nature is governed by cause and effect, and that it is the job of each science to discover causes within its domain. See Norton (2003, 2007) for a description of this view and his argument against it. Ney (2009) appears to think that Norton's *causal fundamentalism* is the same position as her *causal foundationalism*, but the two positions are in fact quite distinct, and they are responses to very different questions. In fact, Jeffrey Sykora has made it clear that four different philosophical positions are available when considering the two orthogonal questions: one can consistently be a fundamentalist/foundationalist, a fundamentalist/anti-foundationalist, an anti-fundamentalist/foundationalist, or an anti-fundamentalist/anti-foundationalist with regard to causation (personal communication, June 1, 2012).

6.2.1 Anti-foundationalist unification

On an anti-foundationalist unification of causation, causal reasoning in physics is horizontal to—*i.e.*, on the same "level" as—the rest of causal reasoning in other sciences and in everyday life. Causal reasoning occurs in the same manner in physics as it does in other branches of life and scientific research. At least as regards causation, there is no foundation-to-edifice relationship. Labeling certain relationships in physics as causal is a way of communicating the same thing that we communicate when we label causal relationships in other contexts—*i.e.*, it is a way of asserting that an external intervention on a certain variable in a certain system, under certain conditions, is associated with change in another variable or variables within that system. It is a kind of shorthand, and inasmuch as philosophers have been able to give a full account of how such assertions work (see Woodward (2003b)), it is not an imprecise shorthand.

But a worry presents itself. If physics turns out to use causal concepts and language in the same manner as other aspects of life and study instead of grounding those practices and modes of reasoning, and if there is no other place where we seem to be able to ground the use of causal content and claims, then what justifies that use? The worry, succinctly stated is this: is all causal reasoning done on the basis of some one (or few) *a priori* principles? Wouldn't such a practice be unacceptable for empirical science?[5]

In chapter 3, I formulated a minimally sufficient and necessary condition under which a variable can be truthfully said to be an interventionist cause of another variable. In a certain sense, it seems that the criterion has an *a priori* status. We seem to declare a prior conviction that we expect to find certain types of patterns in the world: specifically, patterns in the way that one variable responds to interventions on another. This expectation is put into practice when, entering into a scientific investigation, we structure our experiments and studies in a certain way: we use interventions whenever possible, and in so doing we exhibit our interest in certain kinds of patterns. That is, we are interested in finding robust results of a particular sort—not just robust correlations between two *observed* variables, but robust correlations between the observed values of one variable and the *interventions* performed on

[5] This is the worry expressed by Norton (2003) when discussing the "second horn" of the "causal fundamentalist's dilemma".

another.

However, the fact that we do indeed find relationships obeying those expected patterns is an *a posteriori* fact about the world. If the criterion for an interventionist causal relationship identifies and collects together a wide swath of relationships—if there are real relationships in the world that satisfy the criterion—then that is an *a posteriori* fact about the world, despite the fact that there may have been a preconceived notion of a causal relationship in mind prior to the investigation.[6]

So there is a minimal, non-vicious sense in which our choice to ask certain questions about the content of physics—causal questions of the interventionist form that I have characterized in this dissertation—is an *a priori* approach to the phenomena that are the object of physics. However, it is an *a posteriori* question whether or not the relationships discovered in physical experimentation and described in physical theory do or do not meet the criteria for being interventionist causal relationships. By way of analogy to Popper's falsifiability criterion for science, we might say that it is a virtue that our search for causes is "frustratable", in that it may not always find causal relationships where it seeks them.[7]

Moreover, there is a whole field of empirical questions to be asked about causal relationships if and where they are discovered. Woodward (2010) makes the distinction between the question of whether or not a relationship is causal in the "undemanding" sense (*i.e.*, a yes-no question), versus the question of additional conditions that distinguish *among* causal relationships. In many cases, the simple question "Does X cause Y?" has such a low threshold for the "yes" answer that it is not informative enough for our purposes.[8] Experimental science shows its intimate connection to cause-seeking not just in the initial way that it asks, "Does X cause Y?", but also in its continuing search for better answers about the range

[6] An argument along these lines is given by both Reichenbach ([1932] 1978) and Nagel (1961).

[7] In order to preserve the *a posteriori* character of empirical investigation, it is particularly important that we do not commit ourselves to a universal "principle of causality"—*i.e.*, that we avoid the expectation that *every* relationship between any pairing of variables X and Y will be a causal relationship, or that we will find some kind of causal relationship whenever we investigate any observed regularity. It is also worth noting that, on the view sketched here, it becomes an *empirical* question whether or not all relationships that *are* discovered to fulfill the criteria for being interventionist causal relationships share certain properties (*e.g.*, slower-than-light propagation).

[8] Recall just how low the threshold is by looking back at section 3.1.1: For X to be a cause of Y, there must exist *some* intervention I on X with respect to Y, *some* set of background conditions $\mathbf{BC}$ having *some* values, and *some* set of auxiliary variables $\mathbf{S}$ held fixed at *some* set of respective values such that a change is seen in Y.

of background conditions under which a particular causal relationship is "stable" and in its attempt to find the most appropriate level of description of cause and effect so that the description is "proportional" (containing neither too much detail so that the description is misleading, or too little so that the description does not allow for good prediction). Beyond these questions of stability and proportionality, the type of effect exhibited in a particular causal relationship and its mathematical form with respect to the cause variable and time is a question that can only be answered empirically.[9]

6.2.2 Anti-reductionist unification

Another feature of the unified notion of causation obtained by an interventionist analysis of physics is that it is non-reductive. Woodward (2010) has noted some biological cases in which the preservation of non-reductive strategies is important to finding stable and proportionate causal relationships, but one might think that this is an idiosyncrasy of biology. Rather, my thesis has shown that non-reductive/macro-level causal relationships are present in physics.

Seeing the value of higher-level causal relationships requires the interventionist account of causation rather than CVPC. Recall that, on CVPC, a true causal relationship is to be found in the temporal evolution of a closed system for which we have a "complete" state description that allows us to predict any later state in that evolution. The desire for such a "complete" description generally pushes us in a reductive direction: we look for as much detail as possible on the smallest possible level. An advocate of CVPC, when considering thermodynamics, will claim that the true causal relationships are to be found in the evolutions of microstates. In terms of merely predicting the dynamic evolution of a thermodynamic system, a fine-grained state description will prove much more powerful than a macro-scale description in terms of thermodynamic state variables.

The interventionist account of causation, in contrast, acknowledges that relationships between macroscopic thermodynamic state variables can be genuine causal relationships,

[9] An effect variable's "reaction" to an intervention on a cause variable is often characterized simply as a function of the cause variable's value and other relevant factors. However, in some cases, it is important or necessary to model the reaction as dependent on time, and time-dependent effects can take several forms (*e.g.*, change in intercept, change in slope, continuous or discontinuous, immediate or delayed). For an interesting discussion of time-series effects, see Shadish et al. (2002, Ch. 6).

as I showed in chapter 5. The interventionist account does not always push in a reductive direction. Woodward (2010) has discussed some of the virtues by which interventionist causal models can be compared and contrasted. On the one hand, *stability* is one of these virtues; it is a property whereby causal relationships continue to hold across a range of background conditions. The wider the range of background conditions for which a causal relationship holds, the more stable that causal relationship is. Often, our causal relationships become more stable (robust under a wider set of background conditions) when we attend to fine-grained, proximate details. On the other hand, *proportionality* is another virtue in causal relationships that can sometimes lead us to resist reductive description. Proportionality is the property whereby the cause and the effect in a given causal relationship are each characterized in a way that accurately identifies variations in the value of the cause that will lead to the realization of alternate values of the effect. When cause and effect are characterized in a *proportionate* way, these characterizations contain neither too much nor too little detail.

How do considerations of stability and proportionality end up affecting our preferred "level" or "scale" of causal relationships with regard to thermodynamic phenomena? Let's spell the question out a bit more in the case of a specific example. When we considered the pressure-driven process in chapter 5 (see figure 5.5), could we not have described everything in terms of the kinetic theory of gases? After all, is pressure not just the force per area collectively exerted by particle collisions? And is volume not simply a number corresponding to the size of the three-dimensional space currently occupied by a collection of particles? Could we not choose to describe the process in terms of an enormous set of causal interactions constituted by particle collisions and energy transfers through the piston and the walls of the container? The answer, of course, is that we *could* do so. And we could do so while continuing to uphold the interventionist interpretation of the network of causal relationships. For example, we could conceive of tiny interventions in which we change the position or momentum of one or more particles, and we could derive the system behavior resulting from such interventions. However, when it comes to the types of interventions that are actually realizable for humans, we must work with interventions on macroscopic variables such as pressure and temperature. And we choose other variables on the same level for the sake of proportionality.[10] As it turns out, when we describe these interventions on macroscopic thermodynamic variables and the responses of thermodynamic systems to those interventions, these relationships happen to be stable; that is an empirical fact about our world.[11]

Thus, the interventionist account verifies the legitimacy of both micro-scale causal relationships and macro-scale causal relationships. However, the macro-scale relationships are *not* understood to be derivative; they are considered causal in their own right—in virtue of counterfactual responses to interventions—and not in virtue of causal relationships holding at the microscopic level. In this regard, it is interesting to note that, on the interventionist account, CVPC-type relationships may continue to count as causal; a CVPC-type causal relationship merely represents a subclass of interventionist causal relationships in which we have to imagine tiny interventions on the initial state in a process. Therefore, CVPC-style relationships can retain the label of "causal"; however, they do so not in virtue of the features picked out by CVPC, but rather in virtue of counterfactual responses to interventions.

Incidentally, it is a common practice in textbooks about thermodynamics and statistical mechanics to introduce entropy according to its statistical interpretation—*i.e.*, as a property proportional to the probability of a given macroscopic state with respect to all possible (and equally probable) microscopic composition states. Swendsen (2012, xv), for example, considers statistical mechanics to be a "conceptual precursor" to thermodynamics because it predicts macro-scale thermodynamic behavior using a contemporary understanding of atomic structure.[12] This approach betrays a kind of prejudice of scale against macroscopic theorizing—taking such theorizing to be inferior to theorizing at the microscopic level. The

[10] We value proportionality when, for example, we choose to characterize pressure as a cause of volume, as opposed to characterizing pressure as a probabilistic cause of a set of microstates the vast majority of which happen to result in a system volume falling within a certain small range.

[11] The stable macro-scale causal relationships in thermodynamics confirm what some philosophers have said about the futility of blindly looking for more and more detailed descriptions and characterizations. Batterman (2002), for example, emphasizes the necessity of eliminating detail in certain types of physical explanations involving "asymptotic reasoning". Wilson (Forthcoming) argues that Duhem's hostility toward his contemporaries' tendency to hope that the lower-scale molecular realm would provide a way of avoiding dealing with thermal properties was (and continues to be) justified. Mitchell (2009) argues that in complex systems with redundancies, feedback loops, and various ways of reorganizing, the reductionist pursuit of modular causes will often be frustrated.

[12] This also seems to have been both Gibbs' and Boltzmann's view. Gibbs, for example, wrote: "The laws of thermodynamics may easily be obtained from the principles of statistical mechanics, of which they are the incomplete expression..." (Gibbs 1902, ix).

idea is that the macroscopic properties of physical systems are merely a *consequence* of microscopic configurations and behaviors, rather than a domain of study and inquiry in their own right. It seems to me that the same prejudice lies behind a general preference for CVPC over the interventionist view when treating questions of causal relationships in physics. Furthermore, the preference for microscopic explanations leaves one in the curious position of having to acknowledge with some puzzlement, as Swendsen does, that the "conceptual precursor" is the "historical latecomer". The fact that Clausius and others (notably, Thomson) came to the concept of entropy well before our "mature" understanding of atomic structure should not be treated as an oddity; their thought deserves careful attention and study. As I argued in my discussion of Clausius in chapter 5, when we attend to this macro-scale reasoning, we see the original (and persistent) conceptual roots of thermodynamics as well as its interventionist causal structure.

6.3 CONCLUSIONS AND FUTURE RESEARCH

In summary, I have argued for the physical respectability of the interventionist account of causation. There are three specific ways in which interventionist causation is superior to what I have called the "consensus view of physical causation" in describing causal relationships in physics. First, the interventionist account acknowledges the importance of experiment and its epistemic role in testing and gaining knowledge of causal relationships. Second, the interventionist account describes the very type of reasoning that is required by careful theorizing about thermodynamic systems and their interactions. And finally, as I have argued in this chapter, the interventionist account offers the prospect of a unified concept of causation: across theory and experiment, across different sciences, and extending to our everyday concepts.

My arguments have remained confined to thermodynamics and to relatively simple experiments. To more fully flesh out the view in question would require (1) an analysis of more complex modern experiments in physics and the causal inferences they afford; and (2) an examination of the extent to which interventionist analysis is applicable in other areas of theoretical physics. This dissertation has provided the philosophical tools for each of these areas of future research, and as I have explained in this concluding chapter, there seem to be good reasons for hope that these extensions to this project can be accomplished.

BIBLIOGRAPHY

Agassi, Joseph. 2008. *Science and its History: A Reassessment of the Historiography of Science*. Dordrecht: Springer.

Bacon, Francis. (1620) 2000. *The New Organon*. Edited by Lisa Jardine and Michael Silverthorne. Cambridge: Cambridge University Press.

Bain, Alexander. 1870. *Logic*. London: Longmans, Green, Reader, & Dyer.

Barnes, Jonathan, trans. 1993. *Aristotle: Posterior Analytics*. 2nd ed. Oxford: Clarendon Press.

Batterman, Robert W. 2002. *The Devil in the Details*. Oxford: Oxford University Press.

Bell, J. S. (1975) 2004. "The Theory of Local Beables." Chap. 7 in *Speakable and Unspeakable in Quantum Mechanics*, 2nd ed., 52–62. Cambridge: Cambridge University Press.

Boyle, Robert. 1660. *New Experiments Physico-Mechanicall, Touching the Spring of the Air, and its Effects (Made, for the most part, in a New Pneumatical Engine)*. Oxford: H. Hall.

——— 1662. *New Experiment Physico-Mechanical, Touching the Air: Whereunto is added A Defence of the Authors Explication of the Experiments Against the Objections of Franciscus Linus, And, Thomas Hobbes*. 2nd ed. Oxford: H. Hall.

Bridgman, P. W. 1927. *The Logic of Modern Physics*. New York: The MacMillan Company.

Broadie, Sarah. 1982. *Nature, Change, and Agency in Aristotle's Physics: A Philosophical Study*. Oxford: Clarendon Press.

Butterfield, Jeremy. 2007. "Reconsidering Relativistic Causality." *International Studies in the Philosophy of Science* 21 (3): 295–328.

Callen, Herbert B. 1985. *Thermodynamics and an Introduction to Thermostatics*. 2nd ed. New York: John Wiley & Sons.

Cardwell, D. S. L. 1971. *From Watt to Clausius: The Rise of Thermodynamics in the Early Industrial Age*. Ithaca: Cornell University Press.

Carnot, N. L. Sadi. (1824) 1897. *Reflections on the Motive Power of Heat*. 2nd ed. Edited by R. H. Thurston. Translated by R. H. Thurston. New York: John Wiley & Sons.

Cartwright, Nancy. 1983. *How the Laws of Physics Lie.* Oxford: Oxford University Press.

Cioffari, Vincenzo, trans. 1937. "Torricelli's Letters on the Pressure of the Atmosphere." In *The Physical Treatises of Pascal: The Equilibrium of Liquids and the Weight of the Air*, 163–170. New York: Columbia University Press.

Clapeyron, Émile. 1834. "Mémoire sur la Puissance Motrice de la Chaleur." *Journal de L'École Royale Polytechnique* 14 (3): 153–190.

Clausius, Rudolf. (1850) 1851. "On the Moving Force of Heat, and the Laws regarding the Nature of Heat itself which are deducible therefrom." *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science*, 4th ser., 2 (8): 1–21.

——— (1854) 1856. "On a modified Form of the second Fundamental Theorem in the Mechanical Theory of Heat." *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science*, 4th ser., 12 (77): 81–98.

——— 1862a. "On the Application of the Theorem of the Equivalence of Transformations to the Internal Work of a mass of Matter." *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science*, 4th ser., 24 (159): 81–97.

——— 1862b. "On the Application of the Theorem of the Equivalence of Transformations to the Internal Work of a Mass of Matter." *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science*, 4th ser., 24 (160): 201–213.

——— (1865) 1867a. "On Several Convenient Forms of the Fundamental Equations of the Mechanical Theory of Heat." Translated by John Tyndall. In *The Mechanical Theory of Heat, with its Applications to the Steam-Engine and to the Physical Properties of Bodies*, 327–374. London: John Van Voorst.

——— 1867b. *The Mechanical Theory of Heat, with Its Applications to the Steam-Engine and to the Physical Properties of Bodies*. Translated by John Tyndall. London: John Van Voorst.

Clavelin, Maurice. 1974. *The Natural Philosophy of Galileo*. Translated by A. J. Pomerans. Cambridge, MA: The MIT Press.

Cobb, Aaron D. 2012. "Inductivism in Practice: Experiment in John Herschel's Philosophy of Science." *HOPOS: The Journal of the International Society for the History of Philosophy of Science* 2 (1): 21–54.

Collingwood, R. G. (1940) 1998. *An Essay on Metaphysics*. rev. ed. Edited by Teresa Smith. Oxford: Clarendon Press.

Crombie, A. C. 1953. *Robert Grosseteste and the Origins of Experimental Science, 1100–1700*. Oxford: Clarendon Press.

Cropper, William H. 1986. “Rudolf Clausius and the Road to Entropy.” *American Journal of Physics* 54 (12): 1068–1074.

Dati, Carlo. 1663. *Lettera a Filaleti di Timauro Antiate, della Vera Storia della Cicloide, e della Famosissima Esperienza dell’Argento Vivo*. Firenze: L’Insegna della Stella.

de Waard, Cornélius. 1936. *L’Expérience Barométrique: Ses Antécédents et Ses Explications*. Thouars: Imprimerie Nouvelle.

Dear, Peter. 1995. *Discipline & Experience: The Mathematical Way in the Scientific Revolution*. Chicago: University of Chicago Press.

Dill, Ken A. and Sarina Bromberg. 2011. *Molecular Driving Forces: Statistical Thermodynamics in Biology, Chemistry, Physics, and Nanoscience*. 2nd ed. New York: Garland Science.

Dowe, Phil. 1992. “Wesley Salmon’s Process Theory of Causality and the Conserved Quantity Theory.” *Philosophy of Science* 59 (2): 195–216.

——— 2000. *Physical Causation*. Cambridge: Cambridge University Press.

Drake, Stillman. 1970. “Berti, Gasparo.” In *Dictionary of Scientific Biography*, edited by Charles Coulston Gillispie, Vol. 2. New York: Charles Scribner’s Sons.

——— 1973. “Galileo’s Experimental Confirmation of Horizontal Inertia: Unpublished Manuscripts (Galileo Gleanings XXII).” *Isis* 64 (3): 290–305.

——— 1975. “Galileo’s New Science of Motion.” In *Reason, Experiment, and Mysticism in the Scientific Revolution*, edited by M. L. Righini Bonelli and William R. Shea, 131–156. New York: Science History Publications.

——— 1981. *Cause, Experiment, and Science*. Chicago: University of Chicago Press.

Ducheyne, Steffen. 2006. “Galileo’s Interventionist Notion of ‘Cause’.” *Journal of the History of Ideas* 67 (3): 443–464.

Duhem, Pierre. (1914) 1954. *The Aim and Structure of Physical Theory*. 2nd ed. Translated by Philip P. Weiner. Princeton: Princeton University Press.

Earman, John and John Roberts. 1999. “‘Ceteris Paribus’, There Is No Problem of Provisos.” *Synthese* 118 (3): 439–478.

Eberhardt, Frederick. 2007. “Causation and Intervention.” Ph.D. Thesis, Carnegie Mellon University.

Eberhardt, Frederick and Richard Scheines. 2007. "Interventions and Causal Inference." *Philosophy of Science* 74 (5): 981–995.

Fair, David. 1979. "Causation and the Flow of Energy." *Erkenntnis* 14 (3): 219–250.

Field, Hartry. 2003. "Causation in a Physical World." Chap. 14 in *The Oxford Handbook of Metaphysics*, edited by Michael J. Loux and Dean W. Zimmerman, 435–460. Oxford: Oxford University Press.

Finocchiaro, Maurice A. 1989. *The Galileo Affair: A Documentary History.* Berkeley: University of California Press.

Galilei, Galileo. 1612. *Discorso al Serenissimo Don Cosimo II, Gran Duca di Toscana, Intorno alle cose, che Stanno in sù l'acqua; ò in quella si muovono.* Florence: Cosimo Giunti.

——— (1623) 1957. "Excerpts from The Assayer." Translated by Stillman Drake. In *Discoveries and Opinions of Galileo*, edited by Stillman Drake, 229–279. New York: Anchor Books.

——— (1632) 2001. *Dialogue Concerning the Two Chief World Systems: Ptolemaic and Copernican.* Edited by Stephen Jay Gould. Translated by Stillman Drake. New York: The Modern Library.

——— (1638) 1914. *Dialogues Concerning Two New Sciences*. Translated by Henry Crew and Alfonso De Salvio. New York: The MacMillan Company.

——— (1638) 1989. *Two New Sciences: Including Centers of Gravity and Force of Percussion*. 2nd ed. Edited by Stillman Drake. Translated by Stillman Drake. Toronto: Wall & Emerson, Inc.

Gasking, Douglas. 1955. "Causation and Recipes." *Mind*, new ser., 64 (256): 479–487.

Gaukroger, Stephen. 2001. *Francis Bacon and the Transformation of Early-Modern Philosophy.* Cambridge: Cambridge University Press.

Gibbs, J. Willard. 1902. *Elementary Principles in Statistical Mechanics: Developed with Especial Reference to the Rational Foundation of Thermodynamics.* New York: Charles Scribner's Sons.

Glymour, Clark. 1986. "Statistics and Causal Inference: Comment: Statistics and Metaphysics." *Journal of the American Statistical Association* 81 (396): 964–966.

Gyftopoulos, Elias P. and Gian Paolo Beretta. 2005. *Thermodynamics: Foundations and Applications*. reprint ed. Mineola, NY: Dover Publications, Inc.

Haag, Rudolf. 1996. *Local Quantum Physics: Fields, Particles, Algebras.* Berlin: Springer-Verlag.

Harman, Peter H. 1982. *Energy, Force and Matter: The Conceptual Development of Nineteenth-Century Physics*. Cambridge: Cambridge University Press.

Hausman, Daniel M. 1986. "Causation and Experimentation." *American Philosophical Quarterly* 23 (2): 143–154.

Havas, Peter. 1974. "Causality and Relativistic Dynamics." In *Causality and Physical Theories*, edited by William B. Rolnick, Vol. 16 of *AIP Conference Proceedings*, 23–47. New York: American Institute of Physics.

Hawking, S. W. and G. F. R. Ellis. 1973. *The Large-Scale Structure of Space-Time*. Cambridge: Cambridge University Press.

Herschel, John F. W. 1851. *Preliminary Discourse on the Study of Natural Philosophy*. London: Longman, Brown, Green & Longmans.

Hertz, Heinrich. (1893) 1962. *Electric Waves: Being Researches on the Propagation of Electric Action with Finite Velocity Through Space*. Translated by D. E. Jones. New York: Dover Publications, Inc.

Hesslow, Germund. 1976. "Two Notes on the Probabilistic Approach to Causality." *Philosophy of Science* 43 (2): 290–292.

Holland, Paul W. 1986a. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81 (396): 945–960.

——— 1986b. "Statistics and Causal Inference: Rejoinder." *Journal of the American Statistical Association* 81 (396): 968–970.

Kircher, Athanasius. 1650. *Musurgia Universalis Sive Ars Magna Consoni et Dissoni*. Rome: Haeredum Francisci Corbelletti.

Koyré, Alexandre. (1939) 1978. *Galileo Studies*. Translated by John Mepham. New Jersey: Humanities Press.

Lange, Marc. 1993. "Natural Laws and the Problem of Provisos." *Erkenntnis* 38 (2): 233–248.

——— 2000. *Natural Laws in Scientific Practice*. Oxford: Oxford University Press.

——— 2002a. *An Introduction to the Philosophy of Physics: Locality, Fields, Energy, and Mass*. Oxford: Blackwell Publishers.

——— 2002b. "Who's Afraid of Ceteris-Paribus Laws? Or: How I Learned to Stop Worrying and Love Them." *Erkenntnis* 57 (3): 407–423.

Laplace, Pierre-Simon. (1814) 1902. *A Philosophical Essay on Probabilities*. Translated by Frederick Wilson Truscott and Frederick Lincoln Emory. New York: John Wiley & Sons.

Leff, Harvey S. 1996. "Thermodynamic Entropy: The Spreading and Sharing of Energy." *American Journal of Physics* 64 (10): 1261–1271.

Loria, Gino and Giuseppe Vassura, eds. 1919. *Opere di Evangelista Torricelli*, Vol. 3. Faenza: Stabilimento Tipo-Litografico G. Montanari.

Machamer, Peter. 1978. "Galileo and the Causes." In *New Perspectives on Galileo*, edited by R. E. Butts and J. C. Pitt, 161–180. Dordrecht: D. Reidel Publishing Company.

——— 1998. "Galileo's Machines, His Mathematics, and His Experiments." In *The Cambridge Companion to Galileo*, edited by Peter Machamer. Cambridge: Cambridge University Press.

Maignan, Emanuele. 1653. *Cursus Philosophicus Concinnatus Ex Notissimis Cuique Principiis*. Toulouse: Raymundum Bosc.

Malherbe, Michel. 1996. "Bacon's method of science." In *The Cambridge Companion to Bacon*, edited by Markku Peltonen, 75–98. Cambridge: Cambridge University Press.

Martini, Angelo. 1883. *Manuale di Metrologia, ossia Misure, Pesi e Monete in Uso Attualmente e Anticamente Presso Tutti i Popoli*. Torino: Loescher.

Maudlin, Tim. 2011. "Causation." Chap. 5 in *Quantum Non-Locality and Relativity: Metaphysical Intimations of Modern Physics*, 3rd ed., 114–147. Malden, MA: Wiley-Blackwell.

McMullin, Ernan. 1965. "Medieval and Modern Science: Continuity or Discontinuity?" *International Philosophical Quarterly* 5 (1): 103–129.

——— 1978. "The Conception of Science in Galileo's Work." In *New Perspectives on Galileo*, edited by R. E. Butts and J. C. Pitt, 209–257. Dordrecht: D. Reidel Publishing Company.

Menzies, Peter and Huw Price. 1993. "Causation as a Secondary Quality." *The British Journal for the Philosophy of Science* 44 (2): 187–203.

Middleton, W. E. Knowles. 1963. "The Place of Torricelli in the History of the Barometer." *Isis* 54 (1): 11–28.

Mill, John Stuart. (1843) 1973. *A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigaiton.* Edited by J. M. Robson, Vol. 7. Indianapolis: Liberty Fund, Inc.

Mitchell, Sandra D. 2009. *Unsimple Truths: Science, Complexity, and Policy.* Chicago: The University of Chicago Press.

Nagel, Ernest. 1961. *The Structure of Science: Problems in the Logic of Scientific Explanation*. New York: Harcourt, Brace & World, Inc.

Naylor, R. H. 1989. "Galileo's Experimental Discourse." In *The Uses of Experiment: Studies in the Natural Sciences*, edited by David Gooding, Trevor Pinch, and Simon Schaffer, 117–134. Cambridge: Cambridge University Press.

Newman, William R. 2004. *The Art-Nature Debate and the Issue of Experiment*. Chicago: The University of Chicago Press.

Ney, Alyssa. 2009. "Physical Causation and Difference-Making." *The British Journal for the Philosophy of Science* 60 (4): 737–764.

Norton, John D. 2003. "Causation as Folk Science." *Philosopher's Imprint* 3 (4): 1–22.

——— 2007. "Do the Causal Principles of Modern Physics Contradict Causal Anti-Fundamentalism?" Chap. 12 in *Thinking About Causes: From Greek Philosophy to Modern Physics*, edited by Peter Machamer and Gereon Wolters, 222–234. Pittsburgh: University of Pittsburgh Press.

Palmieri, Paolo. 1998. "Re-examining Galileo's Theory of Tides." *Archive for History of Exact Sciences* 53: 223–375.

——— 2008. *Reenacting Galileo's Experiments: Rediscovering the Techniques of Seventeenth-Century Science*. Lampeter: The Edwin Mellen Press.

——— 2009. "A phenomenology of Galileo's experiments with pendulums." *The British Journal for the History of Science* 42 (4): 479–513.

Pascal, Blaise. 1908. *Oeuvres de Blaise Pascal*. Edited by Léon Brunschvicg and Pierre Boutroux, Vol. 2. Paris: Librairie Hachette.

——— 1937. *The Physical Treatises of Pascal: The Equilibrium of Fluids and the Weight of the Mass of the Air*. Translated by I. H. B. Spiers and A. G. H. Spiers. New York: Columbia University Press.

Pearl, Judea. 2009. *Causality: Models, Reasoning, and Inference*. 2nd ed. Cambridge: Cambridge University Press.

Pecquet, John. 1653. *New Anatomical Experiments*. London: Printed by T.W. for Octavian Pulleyn.

Peirce, Charles Sanders. (1898) 1992. "Causation and Force." In *Reasoning and the Logic of Things: The Cambridge Conferences Lectures of 1898*, edited by Kenneth Laine Ketner and Hilary Putnam, 197–217. Cambridge, MA: Harvard University Press.

Pérez-Ramos, Antonio. 1996. "Bacon's forms and the maker's knowledge tradition." In *The Cambridge Companion to Bacon*, edited by Markku Peltonen, 99–120. Cambridge: Cambridge University Press.

Pietroski, Paul and Georges Rey. 1995. "When Other Things Aren't Equal: Saving Ceteris Paribus Laws from Vacuity." *The British Journal for the Philosophy of Science* 46 (1): 81–110.

Power, Henry. 1664. *Experimental Philosophy, In Three Books: Containing New Experiments Microscopical, Mercurial, Magnetical*. London: T. Roycroft.

Redhead, Michael L. G. 1986. "Relativity and Quantum Mechanics—Conflict or Peaceful Coexistence?" *Annals of the New York Academy of Sciences* 480: 14–20.

Reichenbach, Hans. (1932) 1978. "The Principle of Causality and Its Empirical Confirmation." Translated by Maria Reichenbach. In *Selected Writings, 1909-1953*, Vol. 2, 345–371. Dordrecht: D. Reidel Publishing Company.

——— 1956. *The Direction of Time*. Edited by Maria Reichenbach. Mineola, NY: Dover Publications, Inc.

Rolnick, William B. 1974. "Introductory Remarks on Causality." In *Causality and Physical Theories*, edited by William B. Rolnick, Vol. 16 of *AIP Conference Proceedings*, New York, 1. American Institute of Physics.

Rossi, Paolo. 1968. *Francis Bacon: From Magic to Science*. London: Routledge.

Rubin, Donald B. 1986. "Statistics and Causal Inference: Comment: Which Ifs Have Causal Answers." *Journal of the American Statistical Association* 81 (396): 961–962.

Russell, Bertrand. 1912. "On the Notion of Cause." *Proceedings of the Aristotelian Society*, new ser., 13: 1–26.

——— (1948) 2009. *Human Knowledge: Its Scope and Limits*. London: Routledge Classics.

Salmon, Wesley C. 1984. *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.

Schott, Gaspar. (1664) 1687. *Technica Curiosa, Sive Mirabilia Artis*. Sumptibus Wolfgangi Mauritii Endteri, excudebat Jobus Hertz, Typographus Herbipol.

Shadish, William R., Thomas D. Cook, and Donald T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin Company.

Shea, William R. 1977. "Hydrostatics and the Regulative Use of Experiments." Chap. 2 in *Galileo's Intellectual Revolution: Middle Period, 1610–1632*, 14–48. New York: Science History Publications.

Spirtes, Peter, Clark Glymour, and Richard Scheines. 2000. *Causation, Prediction, and Search*. 2nd ed. Cambridge, MA: The MIT Press.

Stapp, Henry P. 1974. "Macrocausality and Its Role in Physical Theories." In *Causality and Physical Theories*, edited by William B. Rolnick, Vol. 16 of *AIP Conference Proceedings*, 87–114. New York: American Institute of Physics.

Swendsen, Robert H. 2012. *An Introduction to Statistical Mechanics and Thermodynamics*. Oxford: Oxford University Press.

Tannery, Paul, Cornélius de Waard, and Armand Beaulieu, eds. 1986. *Correspondance du P. Marin Mersenne*, Vol. 16. Paris: Éditions du Centre National de la Recherche Scientifique.

Thomson, William. 1853. "On the Quantities of Mechanical Energy contained in a Fluid in Different States, as to Temperature and Density (On the Dynamical Theory of Heat, Part V)." *Transactions of the Royal Society of Edinburgh* XX (3): 475–482.

Torretti, Roberto. 1983. *Relativity and Geometry*. Mineola, NY: Dover Publications, Inc.

von Wright, G. H. (1973) 1993. "On the Logic and Epistemology of the Causal Relation." In *Causation*, edited by Ernest Sosa and Michael Tooley, 105–124. Oxford: Oxford University Press.

Wallace, William A. 1972. *Causality and Scientific Explanation*, Vol. 1. Ann Arbor: University of Michigan Press.

——— 1991. *Galileo, the Jesuits and the Medieval Aristotle*. Hampshire: Variorum.

Webster, C. 1965. "The Discovery of Boyle's law, and the Concept of the Elasticity of Air in the Seventeenth Century." *Archive for History of Exact Sciences* 2 (6): 441–502.

Weeks, Sophie. 2007. "Francis Bacon and the Art-Nature Distinction." *Ambix* 54 (2): 101–129.

Whewell, William. 1847. *The Philosophy of the Inductive Sciences, Founded Upon Their History*. 2nd ed., Vol. II. London: John W. Parker.

Wilson, Mark. Forthcoming. "Two Cheers for Anti-Atomism." In *Physics Avoidance and Other Essays*.

Winnie, John A. 1977. "The Causal Theory of Space-time." *Minnesota Studies in the Philosophy of Science* 8: 134–205.

Wisan, Winifred Lovell. 1978. "Galileo's Scientific Method: A Reexamination." In *New Perspectives on Galileo*, edited by R. E. Butts and J. C. Pitt, 1–57. Dordrecht: D. Reidel Publishing Company.

Woodward, James. 2003a. "Experimentation, Causal Inference, and Instrumental Realism." In *The Philosophy of Scientific Experimentation*, edited by Hans Radder, 87–118. Pittsburgh: University of Pittsburgh Press.

——— 2003b. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

——— 2007. “Causation with a Human Face.” Chap. 4 in *Causation, Physics, and the Constitution of Reality: Russell’s Republic Revisited*, edited by Huw Price and Richard Corry, 66–105. Oxford: Clarendon Press.

——— 2010. “Causation in Biology: Stability, Specificity, and the Choice of Levels of Explanation.” *Biology & Philosophy* 25 (3): 287–318.

——— 2013. “Causation and Manipulability.” In *The Stanford Encyclopedia of Philosophy*, Winter 2013 ed., edited by Edward N. Zalta. http://plato.stanford.edu/archives/win2013/entries/causation-mani/.

Yagi, E. 1981. “Analytical Approach to Clausius’s First Memoir on Mechanical Theory of Heat (1850).” *Historia Scientarum* 20: 77–94.

Zhang, Jiji and Peter Spirtes. 2008. “Detection of Unfaithfulness and Robust Causal Inference.” *Minds and Machines* 18 (2): 239–271.

Zucchi, Niccolò. 1648. *Experimenta Vulgata Non Vacuum Probare, sed Plenum et Antiperistasim Instabilire*. Rome: Typographia Ludovici Grignani.

Zwier, Karen R. 2013. “An Epistemology of Causal Inference from Experiment.” *Philosophy of Science* 80 (5): 660–671.