Positron Emission Tomography from 1930 to 1990: The Epistemology and Process of Scientific Instrumentation

Rick Shang
Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/art_sci_etds

Part of the Philosophy of Science Commons

Recommended Citation

This Dissertation is brought to you for free and open access by the Arts & Sciences at Washington University Open Scholarship. It has been accepted for inclusion in Arts & Sciences Electronic Theses and Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.
WASHINGTON UNIVERSITY IN ST.LOUIS

The Department of Philosophy

Dissertation Examination Committee:
Carl F. Craver, Chair
Talia Dan-Cohen
Dennis Des Chene
Allan Hazlett
Ron Mallon

Positron Emission Tomography from 1930 to 1990:
The Epistemology and Process of Scientific Instrumentation
by
Rick Shang

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2021
St. Louis, Missouri
# Table of Contents

List of Figures ........................................................................................................... v  
List of Tables ........................................................................................................... vii  
Acknowledgments ..................................................................................................... viii  
Abstract ..................................................................................................................... xi  

Chapter 1: Introduction ................................................................................................. 1  
  1.1 Positron Emission Tomography ........................................................................ 4  
  1.2 Scientific Instrumentation in Philosophy of Science ........................................ 8  
  1.3 The Use of an Evolutionary Metaphor for the History of Scientific Instrumentation 19  
      1.3.1 The Modularity of Instruments ................................................................. 23  
      1.3.2 Niche Construction .............................................................................. 25  
  1.4 Methods ............................................................................................................ 27  
  1.5 Chapter Summaries ......................................................................................... 29  
  1.6 Conclusion ....................................................................................................... 32  

Chapter 2: The Evolutionary Framework of Scientific Instrumentation .................. 34  
  2.1 Introduction .................................................................................................... 34  
  2.2 The Design Account of Technology Development ........................................... 37  
  2.3 The Evolutionary Framework .......................................................................... 42  
      2.3.1 Modularity ............................................................................................. 48  
      2.3.2 Niche Construction .............................................................................. 52  
      2.3.3 Community Ecology ............................................................................ 54  
  2.4 Conclusion ....................................................................................................... 57  

Chapter 3: 1930s-1950s, The Invention of the Gamma-Ray Detector and the Epistemic Value of Signal-to-Noise Ratio ......................................................... 61  
  3.1 Introduction ..................................................................................................... 61
3.2 The History of Detecting Gamma Rays ................................................................. 64
  3.2.1 Detecting Gamma Rays ........................................................................... 64
  3.2.2 Coincidence Detection and Adoption of Gamma Ray Detection in Medicine  75
3.3 Philosophy, Experiments, and Noise .................................................................. 81
  3.3.1 The Links Between Data and Phenomena ............................................. 82
  3.3.2 A Scientific and Philosophical Account of Signal-to-Noise Ratio .......... 88
3.4 Conclusion ........................................................................................................... 99

Chapter 4: 1950-1970s, The Invention of Positron Camera and the Epistemic Value of
Sensitivity .................................................................................................................... 102
  4.1 Introduction ...................................................................................................... 102
  4.2 The History of Positron Camera ...................................................................... 105
    4.2.1 The Discovery of Phosphorus-32’s Selective Concentration in Tumors..... 106
    4.2.2 Incorporating Phosphorus-32 Into a Tumor Detection System ............. 110
    4.2.3 Positron-Based Detection of Brain Tumors .......................................... 116
    4.2.4 From the First Prototype to the First Positron Camera ......................... 123
  4.3 Toward A Philosophy of Sensitivity .................................................................. 131
    4.3.1 Geometric Efficiency .......................................................................... 137
    4.3.2 Intrinsic Efficiency ............................................................................ 138
    4.3.3 Fraction of Output that Can be Picked Up ......................................... 139
    4.3.4 Absorption and Scattering ............................................................... 140
  4.4 Conclusion ........................................................................................................ 141

and the Epistemic Value of Temporal Resolution .................................................... 144
  5.1 Introduction ...................................................................................................... 144
  5.2 History ............................................................................................................. 146
    5.2.1 X-ray CT .......................................................................................... 148
    5.2.2 SPECT ........................................................................................... 150
    5.2.3 PETT .............................................................................................. 151
  5.3 Understanding Resolution .............................................................................. 162
    5.3.1 Continuing Hacking’s Unfinished Project ......................................... 163
    5.3.2 Resolution ....................................................................................... 166
# List of Figures

<p>| Figure 1.1: | The Eyes of a Gray Drone-Fly | 2 |
| Figure 1.2: | A Block Diagram of the Positron Emission Tomograph Scanner | 5 |
| Figure 1.3: | A Sample Image from Positron Emission Tomography | 7 |
| Figure 3.1: | Increasing signal-to-noise ratios | 62 |
| Figure 3.2: | Example of signal-to-noise ratio | 68 |
| Figure 3.3: | Experiment Data | 69 |
| Figure 3.4: | Scattered Gamma Rays | 76 |
| Figure 3.5: | DeBenedetti et al.’s Qualitative Determination of the Angle | 78 |
| Figure 3.6: | Amplitude SNR | 92 |
| Figure 3.7: | Signal Only | 93 |
| Figure 3.8: | Signal with Mild Noise | 94 |
| Figure 3.9: | Signal with Extreme Noise | 94 |
| Figure 3.10: | Point Spread Function | 96 |
| Figure 4.1: | Percentage of Phosphorus-32 | 108 |
| Figure 4.2: | Experiment Data | 113 |
| Figure 4.3: | Kallmann’s Experiment Data | 115 |
| Figure 4.4: | Hooded Camera | 117 |
| Figure 4.5: | The Prototype from MGH | 121 |
| Figure 4.6: | The Result of a Balanced, Intended Scan | 122 |
| Figure 4.7: | The Result of an Unbalanced Scan | 122 |
| Figure 4.8: | The Second Prototype, Called the Hybrid Scanner | 126 |</p>
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.9</td>
<td>The Ring Design – Two Photomultipliers Uniquely Determine a Crystal…</td>
<td>127</td>
</tr>
<tr>
<td>4.10</td>
<td>PC-1. The First Positron Camera</td>
<td>128</td>
</tr>
<tr>
<td>4.11</td>
<td>The Rate of Oxygen Being Cleared out of the Pulmonary</td>
<td>129</td>
</tr>
<tr>
<td>4.12</td>
<td>Coincidence Detection Mechanism</td>
<td>140</td>
</tr>
<tr>
<td>5.1</td>
<td>Distribution of Journal Categories Publishing PET Research</td>
<td>161</td>
</tr>
<tr>
<td>5.2</td>
<td>Full-Width at Half-Maximum</td>
<td>168</td>
</tr>
<tr>
<td>5.3</td>
<td>Dropping FWHM of PET Scanners from the 60s-70s</td>
<td>170</td>
</tr>
<tr>
<td>5.4</td>
<td>Full-Width at Half-Maximum</td>
<td>172</td>
</tr>
<tr>
<td>6.1</td>
<td>The Comparison of Publications using PET versus fMRI</td>
<td>183</td>
</tr>
<tr>
<td>6.2</td>
<td>Main Path Analysis Results</td>
<td>185</td>
</tr>
<tr>
<td>6.3</td>
<td>Clusters of Publications</td>
<td>187</td>
</tr>
</tbody>
</table>
List of Tables

Table 3.1: Signal versus Noise with and without Coincidence Detection ................ 80
Table 3.2: Bogen and Woodward versus McAllister on Noise and Phenomena ........ 87
Table 5.1: Comparison Table of Imaging Techniques ........................................... 151
Table 5.2: PET’s Pro & Con ............................................................ 153
Acknowledgments

This dissertation is the product of extended team work and collaboration, from members of my department – the Department of Philosophy, from members of my school – Washington University in St. Louis, especially its medical school, from a lot of scholars and friends I have met, interacted with, and learned from at academic events and other social settings, and from the moral support and encouragement of my family and friends.

First of all, I want to acknowledge the support of my dissertation adviser Carl F. Craver. I cannot be thankful enough for his support, both as a scholar and an adviser, and as a friend, who encouraged me to pursue the current research project, dived into all the details with me, found me support across departments and funding agencies, and again and again helped me revise my dissertation to sharpen my philosophical arguments and presentation of history. Carl has generously provided unparalleled support and I cannot fully describe and appreciate what he has offered.

I also want to acknowledge the key inspiration from Edouard Machery. It was in his class and in the ensuing conversations that I came to the realization that the early history of human functional neuroimaging and Positron Emission Tomography is potentially a fascinating History and Philosophy of Science project.

Second, I want to acknowledge the invaluable support from the Digital Humanities Project and a number of philosophy students who contributed to the collection and analysis of data I use in this dissertation. I thank Douglas Knox and Steve Pentecost for their leadership and support at the
Digital Humanities Project. I thank Ronny Bass, Alex DiChristofano, David Gruskin, and Luis Goicouria for their extremely meticulous collaboration on producing the data. I recall so many summer days working on coding and data processing with you all.

I am also extremely indebted to my entire committee. Talia Dan-Cohen, Dennis Des Chene, Allan Hazlett, and Ron Mallon all spent countless hours on giving me critical input to make my dissertation better.

So many other scholars and friends helped me in the process of the dissertation. Thanks to Jerry Cox, Mark Mintun, Michael I. Posner, Marcus Raichle and other participants in the original research on Positron Emission Tomography at Washington University in St. Louis for providing interviews and other access to archival documents. Thanks to David Colaco, Philipp Haueis, Joe McCaffrey, Anya Plutynski, Julia Staffel, and many others for productive conversations on developing the ideas presented in this dissertation. Thanks to Rachel Dunaway, Sue McKinney, Kimberly Mount, Terry Sharp, Dolores Stewart, and many other staff at the Department of Philosophy, the University Library and the Medical School for helping me go through the administrative process and track down pieces of archival documents.

I am also honored to receive the support of Imaging Sciences Pathway Program at Washington University in St. Louis, under the support of T32 Pre-Doctoral funding from NIH.

The journey is not possible without the support of people named above and many others. Thank you all, from the bottom of my heart.

Rick Shang

Washington University in Saint Louis
August 2021
Dedicated to my parents.
Positron Emission Tomography (PET) is a groundbreaking detection system that, among many other applications, enabled neuroscience researchers to detect and image physiological changes in the brain associated with cognitive operations as subjects perform a task. This technology helped researchers learn to explore the physiological basis of human cognition in new ways and was the immediate forerunner of functional Magnetic Resonance Imaging (fMRI), a technology that continues to dominate the neurosciences. PET was also the founding invention of a new field of cognitive science: human functional neuroimaging. Indeed, some would say that PET encouraged a particular way of understanding the brain – through identifying local realizations of cognition and that the ensuing field, neuroimaging, was the key experimental advance in giving rise to cognitive neuroscience properly so-called.

I use PET as a case study of technology-driven scientific change, a form of scientific change exceptionally common in contemporary science and worthy of its analysis, independently of theory-driven scientific change. I am primarily concerned with two questions, one historical and one epistemic:
• What of a general sort can the history of PET teach us about the process of technological change in science?

• What can the history of PET teach us about the norms by which technologies are and should be evaluated?

To answer those questions, I have used historical, anthropological, and bibliometric research methods to reconstruct the biography of this apparatus. I have consulted the original archive of PET research at Washington University in St. Louis, interviewed most of the core researchers involved in the development of PET, and built a database of more than 100,000 articles on PET, fMRI, and neuroimaging that I then analyze to provide quantitative and qualitative information about the state of the field and the acceptance of the technology across time.

As a scaffolding framework for integrating these diverse sources of information, I rely on a core analogy to evolution by natural selection to model different aspects of the causal-historical process of technology-driven scientific change. This stands in contrast to those who understand artifacts and technologies as necessarily products of intentional choices and shaped by their intended purposes (Vermaas et al. 2007; Houkes and Vermaas 2010). These views emphasize the role of an intentional designer and lead us to view the engineering processes as based on the existence and ordering of preconceived and overarching goals. These goals, it is presumed, are broken into more manageable sub-goals, and then realized materially. While this image of technological development is useful as a rational reconstruction, it necessarily neglects the import of serendipity and blind variation, the centrality of material constraints and canalization, and the fluid nature of the purpose for which the technology will come to be adapted.

So, in addition to the intentional model, we should also consider a blinder, more material, more selectionist model – an evolutionary model. Technological development and technology-driven scientific change are heavily constrained, not only by the intentional decisions of designers, but also by material factors, choices of the designer’s descendants and colleagues, and shifting epistemic
norms emerging and receding as a research program takes shape around the technology. The evolutionary metaphor encourages us to acknowledge aspects of this technological history that intention-driven, psychologistic accounts tend to ignore.

Many aspects of a selective process in the organic world have analogies in the evolution of detection technologies. For example, the evolutionary model leads us to consider whether modular structure of technologies enhances their “evolvability,” as modularity of organic structures is thought to do. Modules are structurally and functionally self-contained components with well-defined inputs and outputs to other components. Replacing them with functional equivalents does not impact the function of other components. Because most organisms are modular, they can mutate only certain modules to be more competitive in an environment, without needing to change or eliminate other modules that already perform at least tolerably well. Similarly, PET and technologies in general are modular; PET consists of separable components: a scintillator, a camera, an amplifier, a computer, etc. Because the technology has a modular organization, researchers could improve on one module at a time without disturbing the operation of the other components. In fact, because different modules of PET were developed primarily at different times, the modularity of PET also allows me to tell a chronological history of PET while focusing on one module at a time.

Another example, the evolutionary model leads us to consider how technologies may shape or rewrite the competitive landscape. Evolution by niche construction involves a creature selecting and structuring its environment in ways that increase its fitness. In the case of PET, scientists can try to select and structure an environment in which their apparatus has fertile applications. Like the niche construction process, competing epistemic norms select apparatus, but apparatus is not merely “selected on” by regulating norms. Researchers can actively pick and choose a less competitive environment and shape the relative priorities of epistemic norms governing data generation. I argue that, functioning like “selective forces,” a host of epistemic norms are critical to the development of detection apparatus. They can direct the development of detection apparatus, although their relative
priority may be shaped by the availability of detection apparatus. I further argue that those epistemic norms complement and extend current philosophical discussion on norms of data generation.

I situate my epistemic discussion among contemporary scholars who have shifted attention to scientific practices, especially to data generation practices. Traditionally, when talking about scientific data, philosophers of science such as Hempel (Hempel 1945a, 1945b), Quine (Quine 1951), Carnap (Carnap 1966), Popper (Popper 2002), Goodman (Goodman 1983), and others primarily focused on the warrant relationship between evidence and theory: how evidence guides and should guide the construction, evaluation, justification, and revision of scientific theories. New generations of philosophers, interested in scientific practices, have expanded the attention to the epistemic and practical norms in data generation. Franklin (Allan Franklin 1989, p. 138), Staley (Staley 2004), Chang (Chang 2007, p. 86), Wimsatt (Wimsatt 2007, p. 44), and others for example are primarily interested in whether researchers can obtain data that accurately and reliably reflect target phenomena.

My work advances this discussion by emphasizing norms of data generation besides accuracy and reliability. These norms are distinct from and, I argue, irreducible to concerns about accuracy and reliability. I discuss different norms in each of the substantive chapters. These norms include increasing signal-to-noise ratio, increasing sensitivity, and increasing resolution, among others.

In short, in this dissertation, I use an evolutionary model to offer a causal-historical account of technology-driven scientific change as demonstrated by PET. In the evolutionary model, I especially emphasize that epistemic norms, such as increasing signal-to-noise ratio, are critical “selective forces” that guide the development of technologies. Those norms supplement and extend current philosophical discussions on the epistemic norms in scientific experiments and instrumentation and present a case that epistemic norms beyond and irreducible to accuracy and reliability play significant roles in guiding scientific development. This dissertation further opens up discussion
about driving forces of technological change and the distinct technological and engineering goals, values, and methods reflected in those epistemic norms.
Chapter 1

Introduction

In the late seventeenth century, Hooke published his *Micrographia* and it instantly became a scientific best-seller. In *Micrographia*, Robert Hooke used the most cutting-edge microscope to draw amazingly detailed sketches of the microworld. For one example, shown in Figure 1.1, he described the details of the eyes of a gray drone-fly (Hooke and Jo. Martyn and Ja. Allestry. 1665, p. 175).

As suggested by the title, *Micrographia*, Hooke was interested in describing and recording a world previously invisible to the naked eyes but now observed through new advances of a new detection instrument – the microscope.

Better observation has always played a role in the history of science. Detection instruments move and shape science in the twenty-first century as much as they did in the seventeenth century. Churchland and Sejnowski observe the saying in neuroscience that the field is data rich and theory poor because research in neuroscience is often driven by new technology to generate new data, with much slower developments in theories about brain functions (Churchland and Sejnowski 1992, p. 16).

Positron Emission Tomography (PET) is one such technology that moves and perhaps revolutionizes neuroscience. John Bickle, for example, argues that human functional neuroimaging is one of the
major revolutions in neuroscience (Bickle 2016, p. 2). Other scholars have called it “central” to psychiatry (Shulman 2001), that it has “transformed” cognitive psychology (Eijsden et al. 2009), and that it has caused a “quiet revolution” in neuroscience (Iacoboni 2007). The reason that PET is critical and revolutionary in neuroscience and in its sibling fields is that it enables a novel way to observe and study human physiology and psychology. PET, by introducing radioactive isotopes into the brain, allows researchers to observe physiological changes in the brain non-invasively and in real time. PET is like a microscope, but for the invisible world of real time physiological changes. In practice, it enables neuroscience researchers to detect physiological changes in the brain associated with cognitive operations as subjects perform a task. This technology helps researchers learn to explore the physiological basis of human cognition in new ways and was the immediate forerunner of functional Magnetic Resonance Imaging (fMRI), a technology that continues to shape neuroscience and that quickly found adoption in psychology, other social sciences, and even in the English department (Alexandrov 2007).
Although Positron Emission Tomography has an instrumental role in neuroscience, researchers did not design Positron Emission Tomography for functional neuroimaging, or even for neuroscience specifically. Functional neuroimaging was not even in the vocabulary in the 1950s, when the initial research on PET took place. Instead, in the 1950s, researchers struggled with the post-war reality that their military-related, nuclear research funding was running out. After the war, federal government focused on encouraging commercial and civilian uses of nuclear power. For example, the passage of the Atomic Energy Act in 1954 allowed for patenting atomic technologies, licensing fissible materials, massive declassification of nuclear research, and permission of exporting and sharing of technologies with foreign companies (Creager 2013, p. 138). In 1948, the Atomic Energy Commission established Division of Biology and Medicine to fund research using nuclear power in medicine, especially on cancer treatment (p. 145). The money was clearly turning into a new, civilian direction and researchers had to catch up by finding new civilian uses of nuclear power, especially in medicine and in a commercially competitive way (for other pieces of research telling similar stories, see: Kraft 2009; Rader 2006).

The history of PET is a history of accidental, technology-led change that shaped fields and theories. PET came into being not through the intentional design of an instrument to detect brain functions, but through the zig-zagging path of trial and error, piecemeal revision, and the effort to find and occupy new niches as the science changed around the technology. Researchers started with the need to transform military-related research and ended up with new instruments that enabled entire fields. To better explore and understand this piece of history, in this dissertation I use PET as a case study of technology-driven scientific change, a form of scientific change exceptionally common in contemporary science and worthy of analysis, independently of theory-drive scientific change. I am primarily concerned with two questions, one historical and one epistemic:

- What can the history of PET teach us of a general sort about the process of technological change in science, especially as an instance of technology-driven scientific change?
• What can the history of PET teach us about the norms by which technologies are and should be evaluated?

To answer those questions, I have used historical, anthropological, and bibliometric research methods to reconstruct the lineage of this apparatus and offer something of its biography. In the process, I have made considerable use of the original archive of PET research at Washington University in St. Louis (WUSTL), interviewed most of the core researchers involved in the development of PET, and built a database of more than 100,000 articles on PET, fMRI, and neuroimaging that I then analyze to provide quantitative and qualitative information about the state of the field and the acceptance of the technology across time.

In this introduction, I describe my dissertational project in four sections. First, I introduce my case study, Positron Emission Tomography. Second, I argue that scientific instrumentation deserves more focused and sustained attention in the philosophical literature. Third, I explain the evolutionary metaphor that runs throughout the dissertation. Fourth, I introduce my scholarly methodology. I conclude with a list of chapter summaries.

1.1 Positron Emission Tomography

This dissertation is at its heart a case study in the history and philosophy of neuroscience, focused on the construction of Positron Emission Tomography and its uses in human functional neuroimaging and Cognitive Neuroscience.

Positron Emission Tomography is a detection technology to use the Positron Emission Tomograph (also abbreviated as PET) scanner to produce non-invasive, three-dimensional images of objects, often of organs inside the human body. It works roughly as follows: researchers first procure radioactive isotopes that participate in biological processes of interest to them. They then inject radioactive isotopes into test subjects’ or patients’ bodies. Those isotopes will decay and emit positrons in the location where the desired biological processes take place. The emitted positrons
travel a very short distance before they hit electrons and annihilate. Upon annihilation, one positron-electron pair emits one pair of gamma rays traveling in opposite directions. The PET scanner, the instrument, then picks up the paired gamma rays from the outside. After a series of signal processing steps, the scanner generates an image based on regional radioactive intensity.

Figure 1.2: A Block Diagram of the Positron Emission Tomograph Scanner

Figure 1.2 is one of the early design diagrams for the scanner designed for human brain imaging. The left side is the detection side. Once a pair of gamma rays escaped the skull, they would hit the detection device on opposite sides. To cover all 360 degrees of potentially incoming gamma rays, researchers created six sets of detectors (Banks A-F in the diagram). Each set of detectors contain 11 pairs of NaI – a scintillator, and a Photomultiplier Tube (PMT). The scintillator flashes when gamma rays hit it. This flash is then amplified by the PMT before being captured by a digital camera, which transduces visible light into an electric signal and feeds the signal into the discriminator. The right side of the diagram reveals some of the information processing. A discriminator identifies exactly which scintillator-PMT pair is hit by gamma rays to determine the precise location of gamma ray arrival. Then the discriminator feeds the precise time and location information into the coincidence detector. The coincidence detector registers a dual hit if and only if it finds two gamma rays arriving
at roughly the same time but from opposite detectors. Once the coincidence detector registers a dual hit, it tells the input processor that an annihilation event happens somewhere along the straight line between the opposite detectors. Once the input processor receives enough information regarding annihilation events, it can mathematically infer where the annihilation event takes place and generate an image accordingly.

After information processing, PET scanner generates a spatial map of intensity of annihilation events. Researchers commonly visualize the data of numeric intensity by coloring high intensity with red, low intensity with blue, and projecting the colors on an image of the object or organ under study. One example image from PET, after information process and visualization, is Figure 1.3. The red and yellow areas are areas of heightened radioactive activity when test subjects engage working memory. The blue and green areas are areas of decreased radioactive. Because researchers pick radioactive isotopes that concentrate in places of heightened biological activities, the colored areas roughly correspond to heightened and decreased biological activities.

The history of Positron Emission Tomography is a history of technology-driven scientific change. It both involves a major historical episode of novel, scientific instrumentation and the birth of new fields of research due to the availability of Positron Emission Tomography. It allowed researchers to have images of physiological changes in the brain in real time. Researchers such as Fox (Fox et al. 1986) quickly realized the value of studying human cognition by observing changes in cerebral physiology in real time. Michael Posner, another major figure in cognitive science, quickly joined the team at Washington University in St. Louis and worked with Fox to apply PET to his existing study of the mechanism of attention (Bechtel, Abrahamsen, and Graham 2017, p. 84). The general scientific and social impact of contemporary functional neuroimaging and the particular role of PET – as a detection technology to extract information – provides an illuminating case study of the process by which new techniques are invented, introduced into a field of science, developed
Figure 1.3: A Sample Image from Positron Emission Tomography

through use, and evolved through tinkering and the attempt to occupy epistemic niches that open up as science changes.
1.2 Scientific Instrumentation in Philosophy of Science

In this section, I primarily pick up and situate Hacking’s proposed but unfinished research program in scientific instrumentation. Hacking argues that philosophical debate regarding science, such as realism, requires way more depth than an armchair knowledge of scientific theories (Hacking 1983, pp. 186-187), partially because scientific practice and instruments play significant roles in science. He uses the example of a microscope. Perhaps we all understand the point of a microscope, at least in theory – it magnifies and allows us to see a world otherwise too small to be visible. But if we simply give a microscope to an untrained person, the person will not be able to see anything through it. Using (and designing) the instrument requires a lot more engineering and practical knowledge than the layperson imagines from the armchair. Without the engineering and practical knowledge, armed only with some armchair, theoretical understanding of a microscope, the layperson is unable to genuinely engage with science and understand whether what we see through the microscope is real.

Hacking made this argument way back in the 80s. He could have a particular person in mind: van Fraassen, in Scientific Representation, makes precisely the mistake Hacking talks about. van Fraassen assumes that a theoretical understanding of how the microscope works is good enough to make claims about scientific realism (B. C. v. Fraassen 1980, p. 90). He was a part of the general intellectual revolt in the 60s against thinking that material objects and technologies were primarily derived from scientific theories and merely scientific theories applied in real life (Forman 2007). In the past fifty years, many philosophers responded to his call and developed philosophical insight into scientific and experimental practice (Examples include: Bogen and Woodward 1988; Cartwright 1983; Culp 1994; Allan Franklin 1989; Hudson 1999; Rheinberger 1997). In this chapter, I argue that the philosophical literature on scientific and experimental practices can be strengthened with an overarching perspective on how new technologies are introduced, validated, amended, and extended.
into new research and experimental areas. A primary fruit of this dissertation is a preliminary theory of precisely that sort.

First of all, the general intellectual trend, of which Hacking is a part, to expand the focus of philosophy of science from scientific theories, models, and explanations to scientific and experimental practices is alive and booming (Examples include: Ankeny and Leonelli 2011; Chang 2011; Craver 2007; Sullivan 2018). If we date the origins of contemporary philosophy of science to the pioneering work of Carl Hempel and the logical empiricists of Berlin and Vienna circles and their subsequent diaspora, the philosophy of science has had a dominant emphasis on the development of theories, models, explanations and understanding how these theoretical devices were tested against the world (through evidence, confirmation, etc.) (for a historical introduction, see: Godfrey-Smith 2003). In the early days of the field, comparably little attention was paid to experiments, either to the logic of their design or the way they provide data for the purposes of testing theory.

But as modern philosophy of science developed, philosophers recognized that science is more than just theories. Kuhn, perhaps the founder of History and Philosophy of Science as a modern discipline, is vocal about the significance of scientific instrumentation and experiments. For example, in the 60s, in his definition of a “paradigm,” he notes that a paradigm includes not only laws and theories, but also applications and instruments (Kuhn and Hacking 2012, p. 11). In the 1980s, Hacking famously notes that experiments have a life of their own (for example: Hacking 1983, p. xiii). Echoing Hacking, many philosophers and historians have worked on various ways that experiments can stay independent of developments in theory. Galison points out that traditions of experimental practices can be independent from and parallel to those of theories (Galison 1997). Franklin observes that, in many cases, experimental expertise and practice stay the same in the lab while theories change (Allan Franklin 1997).

The development of philosophical interest in scientific instrumentation is accompanied by philosophically interesting questions. Perhaps among the most systematic is Franklin’s research program
on the justification of experimental results. Franklin is curious about how experiments give us data and, more importantly, how researchers are justified to believe that experimental results are of the real world (instead of being an artifact from the apparatus, misreading, or misinterpretation) (Allan Franklin 1989, p. 3; Franklin and Perovic 2015) (for similar themes, also see Chang and Leonelli 2005a; B. C. v. Fraassen 2008; Tal 2015). In other words, he focuses on the norms of accuracy and whether instruments reveal the world to us as it is.

In a similar fashion, philosophers have emphasized instruments as critical in scientific inquiry. Pickering, for example, also notes that science is a performance navigating between human and material agency (Pickering 1995, p. 22): scientific inquiries involve mutual adjustments from experimenter, apparatus, and the natural world.

However, despite contemporary developments in philosophy of science, many of Hacking’s identified challenges in scientific instrumentation deserve more philosophical attention. Here, I am quoting from his 1981 paper, Do We See Through a Microscope? (Hacking 1981) which is also included in his essay collection, Representing and Intervening (Hacking 1983), published in 1983. In this paper, Hacking suggests that understanding how physicists come to understand the world and wading into philosophical debates such as realism require more than an armchair knowledge of how science is done (pp. 186-187). Even the light microscope is “a marvel of marvels,” and does not work in ways that “untutored people suppose.” (p. 186) Hacking goes on to talk about the history of, mechanism of, justifications of, and required expertise for using microscopes. To better help philosophers understand microscopes, Hacking makes the following observations, or four kinds of challenges facing scientific instruments:

- Technical Challenge: Coming up with a new detection instrument is often a distinct technical challenge. We do not see “a world within the world” just by sliding a lens between our eyes and a drop of water. While Leeuwenhoek pioneered the use of the microscope in the 17th century, the microscope mostly remained an entertaining toy during that time. Early
microscopes suffered from optical aberrations. Looking through the microscope used to be similar to looking through a kaleidoscope. It took significant improvement over lens crafting, optical theories, and engineering improvements for the microscope to become useful.

- Normative Challenge: Engineering often follows its own norms. Researchers working on improving the microscope followed a distinct set of norms, not fully overlapping with and often crosscutting theoretical norms. For example, Hacking mentions that researchers intended to increase resolution. He describes resolution as the resolving power: the extent to which we can tell two small dots apart. Following the norm of increasing resolution then is to find ways to tell smaller and smaller dots apart.

- Social Challenge: Social and scientific acceptance is not an automatic process. Despite the better microscope from Zeiss – the same German company we know today – it faced resistance for a long time. Hacking lists two reasons. First, traditional craftsmanship peaked at the same time and competed against Zeiss. Second, Zeiss faced the long-standing suspicion that only the unaided, naked eyes saw the world as it was (A lingering suspicion that one can identify in van Fraassen’s writing. B. C. v. Fraassen 1980, p. 16). The microscope, it was thought, surely created some kind of illusion. A technically superior instrument does not automatically translate into acceptance and adoption.

- Personal Challenge: Using a new detection instrument often requires a special kind of expertise. We do not merely look into the microscope and see the microworld as we hope to. Instead, we need to prepare the specimens, mount them, adjust both the microscope and the specimens, and then look “in the right way.” Using the seemingly simple instrument is not straightforward or intuitive.

Reviewing some examples in the philosophical literature demonstrates that while many philosophers of science are approaching, demonstrating, and responding to those challenges in scientific instrumentation, there is substantial space for additional philosophical discussion, especially on the
interaction between technical, normative, and social challenges of scientific instrumentation. I begin my examples with Chang’s work on the construction and epistemology of instruments in his classic *Inventing Temperature* (2004) and point out that philosophers should follow his example while also turning attention to the technical challenges in the 20th century. Chang’s book is focused on one central question: how did researchers develop increasingly precise and reliable thermometers when the concept of temperature was still in its infancy and remained elusive? He describes the history of iterative improvements both on the instrument and on the concept of temperature. Ultimately, he suggests that the history embodies a form of coherentism (Chang 2007, p. 220):

> The only productive way of dealing with that circularity is to accept it and admit that justification in empirical science has to be coherentist. Within such coherentism, epistemic iteration provides an effective method of scientific progress, resulting in the enrichment and self-correction of the initially affirmed system. This mode of scientific progress embraces both conservatism and pluralism at once.

Here, Chang addresses, among others, a chicken-and-egg problem: to evaluate our measurement of a phenomenon partially depends on our theoretical understanding of what this phenomenon is and does; but our theoretical understanding of what this phenomenon is and does often depends on our measuring it. In the quote, Chang suggests that to break out of the chicken-and-egg problem, we have to subscribe to a coherentist framework. In the framework, researchers begin with a certain stable, coherentist framework of knowledge. For example, researchers can build a thermometer that roughly corresponds to our sense of heat and chill, but may freeze below a certain point, or the temperature readings might not be linear to temperature changes. Researchers then iteratively refine both the instrument and our understanding of temperature. The iterative stages can include empirical improvements, but also theoretical and methodological improvements.

The challenge with Chang’s work is that contemporary researchers are often not troubled by the epistemic concerns of the 18th or 19th-century physicists and chemists. Many contemporary
philosophers of science still rely on cases such as the microscope and the telescope (such as van Fraassen B. C. v. Fraassen 2008). It is not that those concerns do not matter at all to contemporary researchers. In some fields, such as basic physics, researchers still face the challenge that their new measurement instruments and their fundamental physical concepts are equally elusive. But in many other fields, contemporary researchers often work on different sets of key concepts, presuppositions, and research realms. In the field of neuroscience, researchers commonly do not struggle with definitions of time, distance, or the physical definition of X-rays. They take those basic physical phenomena for granted. Their concerns include using established basic physical theories and concepts and their engineering ingenuity to create access to new kinds of knowledge. In the case of PET, researchers do not wonder if positrons or gamma rays are real. They do not wonder if we can have a good measurement to measure the intensity of gamma rays. Instead, they use nuclear research surrounding gamma rays in the 40s to build an instrument to detect changes inside the body non-invasively and are not too bothered if such observed changes are real. So whereas Chang has helped us investigate the technical challenges of and the epistemic relationship between phenomena and their measurement in the 18th or 19th century, I extend his work and pay more attention to the technical challenges and the epistemology of instruments in the 20th century and in new fields that did not exist in the 18th century. Those sciences and fields (such as neuroscience) can borrow theoretical postulates in the process of building and justifying their instruments and have their own challenges.

Consider next Franklin’s work on instrumentation. Franklin pays primary attention to the justification of knowledge produced by instruments and experimentation. Franklin has a similar complaint to mine. He suggests that philosophers of science are not paying sufficient attention to scientific experiments. His 1986 book, after all, is titled The Neglect of Experiment. He also identifies scientific instruments as indispensable in scientific experimentation. But his focus is fairly narrow. He is primarily interested in ways we are justified to believe the results that instruments deliver (Allan Franklin 1989, p. 3). He focuses specifically on calibration. Calibration is the process to standardize
an instrument and a strategy to establish that our data are valid (A. Franklin 1998, p. 237). Here, Franklin has an interesting definition of valid data. Data do not need to be correct or true to be valid. Data are valid when they are of the target phenomenon. He is primarily concerned with whether our instruments or our use of our instruments creates artifacts not existing in the phenomenon in itself (Allan Franklin 1990, p. 104).

Franklin’s work on instrumentation could be further strengthened by deepening the historical aspect of instrumentation. I do not mean that Franklin does not care about the history of tool development. I mean that a much longer historical dimension, with focus on the technical and social challenges of instrumentation, could better explain the pairing of instrumentation and data: why do researchers think instrument X is a good tool to generate data of phenomenon Y? In particular, researchers often need to invest years if not decades on improving an instrument and then spend time during an experiment to calibrate the instrument, before it can generate good data. How do researchers make the decision, before development and calibration, that X is good for measuring Y? Franklin in his 1998 book, Can That be Right? uses a case where physicists at Princeton University used a Cerenkov counter to detect positrons for their experiments on the violation of the combined particle-antiparticle and space-reflection (CP) symmetry (A. Franklin 1998, pp. 239-240). But why should physicists use a Cerenkov counter to detect CP violation? Why did they choose to painstakingly calibrate a Cerenkov counter, when it was not giving out valid data initially, instead of using something else? Chang has already given an answer: the thermometer is a good tool to measure temperature, partially because researchers developed the thermometer in order to measure temperature and, in the ensuing years, the iterative improvement of the tool was tied to the iterative improvement of the theoretical understanding of temperature. In other words, Chang says the pairing between the instrument and the concept co-evolved and “bonded up” in history. In this dissertation, I supplement Chang’s answer with an additional explanation: many times researchers use technology X because they have been working with and are invested in technology X. They have a hammer and they are looking for nails – or academic questions.
Next and closer to the home theme of this dissertation, consider Bickle’s discussion of neuroscientific revolutions. Bickle, in response to Kuhn, explicitly suggests that tool development drives paradigm shifts and revolutions in neuroscience. Bickle notes, along with many other philosophers, that although Kuhn talks about experiments and instruments, his theory of revolutions is still focused on scientific theories and models. According to Kuhn, paradigm shifts and revolutions take place because of theoretical conflicts, anomalies, etc. But in neuroscience, revolutions often come from new tools (Bickle 2016, p. 2). One example that Bickle uses is optogenetics. Optogenetics is a research field enabled by a critical technique: highly precise and selective control of the activation or inhibition of target neurons (p. 9). This technique opens up a new venue to explore casual-mechanistic hypotheses of brain functions (p. 10). With the technique, researchers publish hundreds of papers each year to demonstrate the downstream effects of selectively manipulating a few neurons in animals’ brains.

The challenge with Bickle’s work is that, although he centers the role of instruments in enabling scientific revolutions, he simplifies the process of how instruments drive scientific change. Supplementing Bickle, I point out that the process of inventing and addressing challenges of instrumentation is intertwined with scientific change. Researchers do not suddenly have, for example, a completed, modern optogenetics system and started to create a revolution. Instead, the revolution and development of new methods and new scientific directions to use optogenetics is intertwined with the historical, inventive process in which researchers slowly and gradually developed, adjusted, and perfected optogenetics tools. The “revolutions” that Bickle alludes to are deeply tied to the historical dimension and the technical, normative, social, and personal challenges of new instruments.

While scholars like Chang, Franklin, and Bickle have all center the role of instruments in their thinking about science, they have not fully addressed or responded to the four challenges of scientific instrumentation raised by Hacking. Understanding how researchers respond to those four challenges is not merely a historical curiosity. Understanding the process enables philosophers to understand
how researchers invent novel scientific instruments and how the invention may shape or even drive scientific change. In this dissertation, I reconstruct how researchers on PET faced and largely conquered the four challenges and extend my observation into a philosophical thesis. While I cover all four challenges, I pay special attention to the technical, social, and normative challenges. On the technical and social side, I agree with Hacking that the technical invention and adoption of a new instrument is a complicated process. It often involves significant modification of both the instrument and the related context of scientific inquiries for an instrument to find its lasting role in the sciences. I recount how researchers working on prototypes of Positron Emission Tomography zigzagged and changed their nascent instruments’ features and use-cases for well over a decade before they saw widespread acceptance of Positron Emission Tomography. The process is not merely a process of identifying and then appealing to existing interests in medical research. Instead, very often, researchers develop new problem domains for their technologies.

On the normative side, I systematically explore the changing and unique landscape of epistemic and engineering norms in instrumentation over the course of PET’s lifetime through the 1990s. I situate my epistemic discussion among contemporary scholars who have shifted attention to scientific practices, especially to data generation practices. Traditionally, when talking about scientific data, philosophers of science such as Hempel (Hempel 1945a, 1945b), Quine (Quine 1951), Carnap (Carnap 1966), Popper (Popper 2002), Goodman (Goodman 1983), and others primarily focused on the warrant relationship between evidence and theory: how evidence guides and should guide the construction, evaluation, justification, and revision of scientific theories. New generations of philosophers, interested in scientific practices, have expanded the attention to the epistemic and practical norms in data generation. Franklin (Allan Franklin 1989, p. 138), Staley (Staley 2004), Chang (Chang 2007, p. 86), Wimsatt (Wimsatt 2007, p. 44), and others for example are primarily interested in whether researchers can obtain data that accurately and reliably reflect target phenomena.
My work advances contemporary philosophers’ work by introducing a new angle to understand the normative challenges researchers face: the normative challenges of engineering instruments. In Franklin, Staley, Change, Wimsatt, and other philosophers’ discussion, researchers care primarily about the relationship between data and target phenomena – such as if data accurately reflect target phenomena. While researchers no doubt do care about the quality of data, there are many scenarios in which researchers rely on the relationship between scientific instruments and target phenomena: data are good and usable because the instruments are in the right causal and normative relationship with target phenomena. One can dig into Franklin’s discussion and apply it in my thesis: researchers want to calibrate their instruments because they want to use known, controlled target phenomena to ensure instruments are in the right kind of causal and normative relationship with future, unknown target phenomena. So when instruments start to generate data about unknown, target phenomena, researchers know they can trust the data are of the phenomena.

My discussion in a way is similar to Plato’s question in Euthyphro: whether God loves the pious because it is pious, or it is pious because God loves it. In my discussion, I point out both are the case in science. Researchers in many scenarios care about the relationship between data and target phenomena. But in many cases, directly verifying the relationship is time-consuming, difficult, otherwise undesirable, or simply unnecessary. Researchers trust the data because the data are generated by the right kind of instruments. The scientific practice of calibration and validation is such a good example. For example, when we are using a weighing scale, we often use a known weight to ensure the scale works correctly. Then we trust the scale to be good for quite a while, before we have reasons to re-adjust and calibrate (such as after the scale is in the storage for a long time).

But what makes instruments good and trustworthy? Perhaps the most immediate answer is that instruments are good and trustworthy because they generate accurate and precise data. Instruments just are a means to our end of getting the data anyways. Throughout the dissertation, I argue this
picture is incomplete because unlike data, instruments are causal mechanisms. People do not trust instruments only because they, when being verified, generate good data. People trust instruments also because, in addition to generate good data, we have a good causal understanding of what they do, and when they do not generate good data, there are intervening methods to make them so. Let us think about the calibration process again: scientific calibration is not merely verifying if an instrument works. Calibration involves sometimes painstaking process to adjust and eventually make an instrument work (see Franklin A. Franklin 1998, p. 237). Causal understanding and expected causal intervention are integral parts of the design and evaluation of instruments. To better understand what kind of causal understanding and intervention are involved in the evaluation of instruments, I rely on the history of PET to argue that those epistemic norms share at least three core features:

- Those norms are uniquely of the instruments – material, causal mechanisms, not of the data
- They measure aspects of those causal mechanisms
- They offer clear, mechanistic information to improve the function and performance of those instruments

In short, my dissertation fleshes out contemporary philosophers of science’s focus on scientific instruments and scientific practice and responds to Hacking’s identified challenges in scientific instrumentation. In particular, I articulate the normative challenges in scientific instrumentation by providing systematic accounts of epistemic and engineering norms of scientific instruments. But so far, I still have not discussed how researchers address the technical and social challenges of instrumentation and how addressing those challenges can introduce scientific change. In the next section, I argue the process has a striking resemblance to an evolutionary process.
1.3 The Use of an Evolutionary Metaphor for the History of Scientific Instrumentation

In this section, I pick up the discussion in the previous section and motivate the use of a specific evolutionary metaphor to understand the history and associated social challenges in scientific instrumentation. I argue that an evolutionary metaphor is extremely useful for understanding how researchers create an instrument from within their own social and theoretical context. I emphasize how they use the instrument to interact with that social and theoretical context and push for scientific change.

Philosophers of science since the days of logical positivism are famously interested in the context of justification more than the context of discovery. I do not mean that the context of discovery does not receive attention by philosophers (Examples include: Darden 1991; Nickles 2006). I do mean that the lingering Popperian influence persists in the background: the context of discovery is too irrational, too messy to properly belong to philosophy (Popper 2002, p. 8). Maybe psychology or the social sciences in general are more fitting homes for the context of discovery. At the same time, an a-rational and fairly “messy” process has received enormous attention by philosophers of science: the evolutionary process. I argue that the study of and insight from the evolutionary process can produce philosophical insights in the context of discovery, especially of how technologies are discovered and improved. The evolutionary process does not provide an algorithm for discovery, but it does illuminate how social, technical, and epistemic concerns arise and interact with each other along the way.

The history of PET bears an immediate, prima facie similarity to the process of evolution by natural selection. The history of PET is a history of accidental, bottom-up invention. Researchers intended to save their existing nuclear research projects, when World War II came to a close. Very few, if any, would envision that their unorganized and uncoordinated efforts to find civilian applications of
their nuclear research ended up in a medical imaging technology. It was the endless searching and zigzagging of researchers, under severe restrictions of available materials and shifting institutional resources, power, and other social pressures that ended up giving us PET. No single person produced a blueprint for PET at the beginning.

The analogy looks more substantial as we dig into the similarity. Here, I primarily adopt the framework for understanding evolution by natural selection proposed by Hull, who extends and applies the evolution by natural selection model to the development and transmission of ideas, concepts, and theories. Hull proposes an abstract evolutionary model that can be used to describe biological, social, and conceptual developments (Hull 1988b, p. 13). He identifies four key, abstract components and processes that can be instantiated and applied to processes outside biological processes (Hull 1988a, p. 134):

“a process in which the differential extinction and proliferation or interactors cause the differential perpetuation of the replicators that produced them.” (Italics original)

Here, in the quotation, Hull suggests that a type level abstraction of evolution should include:

- A replicating population
- that can in the long run (“perpetuation”)
- benefit from or be harmed by (“extinction” and “proliferation”)
- causal interactions (“cause”) with other features in the world

In other words, Hull suggests the following components and processes as central to an evolutionary framework:

- Unit of Selection: a replicating population of varying units that have heritable features to be selected on
- Fitness Currency: there is a criterion or many criteria to benefit or harm the population
• Causal Interaction: there has to be causal interaction between the population and environmental factors

• Long Term Effect: the causal interaction should, through the fitness currency, lead to long term differential results on units of selection

In addition to identifying the type level abstraction, Hull suggests that an evolutionary process is an accurate description of the development of scientific ideas and theories (broadly construed) (Hull 1988a; 1988b, p. 9). Hull identifies the relevant key components and processes of evolution in scientific development.

• Unit of Selection: the population of ideas (Hull 1988a, p. 141). Although ideas exist in scientists’ mind, they do exist for a long time and ideas give rise to ideas – or that they “mutate” and “reproduce” through scientists’ efforts.

• Fitness Currency: Ideas benefit or are harmed through the crediting process in science (p. 129). The most harmful thing we can do to ideas is to ignore them. And ideas “benefit” in the sense that they are cited and integrated into later ideas.

• Causal Interaction: Scientists interact among themselves and with the natural world to generate, improve on, and sometimes abandon ideas (p. 141).

• Long Term Effect: Ideas can be transmitted or forgotten in the long run (p. 140).

Hull’s application of evolution by natural selection on ideas runs into many theoretical challenges. Rosenberg’s critique of Hull identifies perhaps two of the most serious obstacles (Rosenberg 1992). First, abstract entities such as ideas and theories are types, but evolution takes place at the token level. Second, it is very challenging to demonstrate how abstract entities have causes and effects.

Scientific instrumentation potentially fits the evolutionary framework better than scientific ideas, precisely because in this domain, we easily avoid the aforementioned obstacles.
• Unit of Selection: each instrument and group of instruments. Instruments “inherit” in the sense that researchers most often build instruments out of existing components and as improvements or deviations from other existing instruments. Instruments “mutate” because researchers gradually modify their designs.

• Fitness Currency: instruments “benefit” or “are harmed” by adoption and replication. Like ideas, the worst treatment to any instrument is to ignore it and the most ideal fate is to be widely used, standardized, and deployed in a variety of research scenarios.

• Causal Interaction: Researchers can use the instruments to interact with the natural world – such as making observations, speeding up experiments, that enables, increases, and sometimes decreases their adoption and usage.

• Long Term Effect: Modifications of instruments could lead to their long-term popularity or cause them to disappear from current usage.

In the list of features, unlike ideas, each individual instrument is a token that can be subject to selective forces. Second, material, causal instruments are indeed causally efficacious. One of the primary problems with applying the evolutionary metaphor to ideas and theories is that they are abstract, type-level entities that do not causally interact with the world. Instruments are causally efficacious – they are causal mechanisms themselves and can causally interact with the world. In addition, unlike ideas, those material instruments have parts and blueprints – they can be tinkered, modified, replicated, selected, get adopted, and/or lose their places in the scientific world. The evolutionary analogy is much more apt for material instruments than for ideas and theories.

One might complain that the evolutionary metaphor is facile and can be made to apply to many complex social phenomena, while offering little explanatory enlightenment. However, in my dissertation, I argue that the metaphor not only fits the history, but also leads us to see new insights in the development of technology – especially the interaction between technical development and
the social context – we would not otherwise be able to see. In particular, I highlight how theories of modularity and niche construction in evolution by natural selection can offer insights to understand the scientific instrumentation process.

1.3.1 The Modularity of Instruments

One major difference between ideas and instruments is that the latter, as a category, are modular. Instruments almost always have parts and those parts have semi-independent functional roles. I do not mean ideas are not sometimes modular (Quine 1956, Breaking down ideas to first order logic is familiar to any student of philosophy of language. For example, see:). But it is a challenge to identify how we can break down the concept of “yellowness” into stable, structural parts while each part having its functional role. Having parts and being modular render instruments a much better candidate for the evolutionary metaphor for at least two reasons. First, instruments are structurally similar to biological organisms. Both have parts and are modular. Second, the development of scientific instruments is more similar to the evolutionary process. Researchers develop scientific instruments often not by starting from a blank slate, but by tinkering and modifying parts in existing instruments. At the same time, the use of the evolutionary metaphor allows us to generate new insights into the scientific instrumentation process. In particular, I argue that modularity allows researchers to quickly devise and improve on their instruments without a deep understanding of the high-level theories.

Modularity, in essence, is a feature of complex systems, which can be broken down into structural and functional sub-systems and components. Following Simon’s work (Simon 1962, p. 468), contemporary biologists propose the following four characteristics for modularity (Callebaut and Rasskin-Gutman 2005, p. xv):

- Hierarchical Organization. Modularity exists in a system that can be decomposed into sub-systems that span a hierarchy of parts and components.
• Near-Decomposability. “Modules” can be reasonably cleanly decomposed, i.e. isolated from one another through reasonably intelligible and limited interfaces, and replaced with functional equivalents.

• Quasi-Functional Independence. “Modules” should more or less be self-contained functional units. It is understood that modules do need input and output from, and sometimes other dependence on other modules. Therefore, independence is not a binary concept but a matter of degrees. Modules mostly have only quasi-, instead of full, independence.

• Recurrence. Research in homology suggests that modules are often recurring (Müller and Newman 2003). Biologists can often identify similar functional units across a very wide spectrum of animals.

Modularity is fundamental to evolution because modularity enhances evolvability. The short story, adapted from John H. Holland’s theory (Holland 1992), is that a modular system does not need to go through wholesale redesign to improve itself. Organisms can evolve to have, for example, longer tongues to access food without impact on their heart function. Similarly, researchers can retain the working modules, switch out the harmful modules, and test out the new instrument. Replacing the bad modules does no or does minimal harm to the working modules. In comparison, researchers working with a non-modular system would need to throw away what works and go through wholesale re-design if they want to get rid of what does not work.

The evolvability point seems to apply perfectly well to instruments. Instruments “benefit” or are “harmed” insofar as they are used or ignored by other researchers. So researchers want to find ways to improve on their instruments to improve on adoption and reduce elements that turn people off. Instruments are indeed modular too in the sense that we can retain the working modules and replace the non-working modules. Positron Emission Tomography is modular itself. In Figure 1.2, Positron Emission Tomography consists of structurally and functionally separable components: a scintillator, a camera, an amplifier, a computer, etc. Because the technology has a modular
organization, researchers could improve on one module at a time without disturbing the operations of the other components.

In the dissertation, I argue that modularity is a key success factor for Positron Emission Tomography. Because of modularity, researchers were able to swiftly adopt new developments in nuclear research into existing medical instrument designs. Also because of modularity, researchers were able to come up with instrument designs that may potentially work in a wide variety of medical disciplines and quickly identify potential venues for adoption. In fact, because different modules of Positron Emission Tomography were developed primarily at different times, the modularity of Positron Emission Tomography also allows me to tell a chronological history of Positron Emission Tomography while focusing on one module at a time.

1.3.2 Niche Construction

In addition to modularity, niche construction is another example of the way the evolutionary metaphor offers new insights into the scientific instrumentation process. In particular, in ways very similar to the evolutionary process, researchers rewrote the priorities of governing norms in scientific instrumentation to confer a competitive advantage on PET.

Niche construction theory is a theoretical development to challenge the old idea that organisms are passive in natural selection, and so merely react to and are selected by the environment over which they have little direct control. Niche construction theorists point out that organisms actively choose or reconfigure their environment to their own benefit. These reconfiguring activities are then, themselves, replicated in selective processes. For example, beavers build dams to create ponds and to protect them from predators; spiders build webs to change the environment to capture preys.

Odling-Smee, in the 80s and following Lewontin and Waddington, was one of the first biologists to propose a systematic account of niche construction. The niche, according to Odling-smee et al., is “the sum of all the natural selection pressure to which the population is exposed” (Odling-Smee,
Niche construction, then, occurs when an organism modifies the relationship between parts of the organism and parts of the environment by actively changing parts of the environment, by perturbing the environment, or by relocating to a more favorable environment (Odling-Smee, Laland, and Feldman 1996, p. 41).

Niche construction is beneficial for organisms because it changes the evolutionary landscape. According to niche construction theory, organisms do not always search for answers for existing questions. They can rewrite the questions so they, and to the disadvantage of others, already have the answers.

Niche construction theory is useful in understanding tool development because it emphasizes researchers’ role in reshaping the competitive landscape and the selective forces at play into which their instruments are received. It is a model to describe and explain how scientific instrumentation can shape the context of scientific inquiries. Instead of merely appealing to other potential users of a new tool, researchers can change other users’ preferences, nudge the direction of scientific inquiries, or tap into underserved populations, and find new niches to exploit for the adoption for their new tool. Niche construction theorists point out that, by successfully shaping the competitive landscape, organisms often leave long term impacts on the general environment. The environment is heterogeneous and fluctuates all the time. As organisms construct their niches in the environment, the selective pressure changes accordingly, and what counts as adaptation will fluctuate as well (Odling-Smee 1988, p. 90). For example, birds probably first evolved feathers for thermoregulation. But once feathers were exaptated for flights and changed the competitive landscape, other body parts (such as chest muscle) went through secondary evolution to better facilitate flight (Gould and Vrba 1982, p. 11).

We observe a similar phenomenon in the history of scientific instrumentation. Of course, researchers do not face predators, at least not literally. Instead, I argue that a host of epistemic and engineering norms serve a similar role like “selective forces” and are critical to the development of scientific
instruments. Those norms can direct the development of detection apparatus such as PET, although their relative priority may be shaped by the availability, use, and popularity of existing scientific instruments. Hence researchers could rewrite the competitive landscape and change the relative priorities of those epistemic and engineering norms to accelerate the adoption of their own instruments. For example, researchers at WUSTL found a relatively untapped “environment”, an “environment” selected by the norm of increasing temporal resolution (colloquially known as “speed”). Being able to witness cerebral physiological changes allowed researchers to understand how regional cerebral physiological changes correspond to human cognition in an entirely new way. Previously, lesion studies were the dominant method to understand how human cognition was realized in cerebral physiology. But lesion studies require lesions that researchers cannot precisely or ethically control. And while lesions help to establish the relationship between lesions and cognitive functions, they do not allow for real time observation of how other regions interact with each other or are involved in daily cognition. Researchers at WUSTL quickly realized that there is an unoccupied “niche” – the niche to observe cerebral activities, both abnormal and normal activities in real time and adapted their instrument to occupy the niche. And their success shaped the landscape so that multiple imaging techniques started to compete more seriously in the niche.

1.4 Methods

To support my arguments for engineering norms and an evolution-like process in scientific instrumentation, I rely on a historically grounded, philosophically motivated case study of the history of Positron Emission Tomography. I approach the history from at least three angles: from a structural perspective, from the perspective of a stranger, and from the perspective of a member of the field. Every historian needs to situate him- or herself between two extremes (Shapin and Schaffer 1986, pp. 4-5): as a deeply involved community member, or a complete stranger looking from the outside. As a member of the scientific community one studies, one has enormous advantage over an untrained stranger in understanding the technical details and decision-making procedures of the scientific
community. But as a member, one often has been so acculturated into certain practices and beliefs that they are taken for granted. Some practices and beliefs that are puzzling or peculiar to a stranger can seem self-evident to a member of the scientific community.

I am enormously fortunate to be able to strike a balance between the two extremes. On the one hand, my doctoral program and my fellowship at the medical school required me to go through the same technical training in functional neuroimaging as an MD or PhD in neuroscience would go through in their first year. It means that I have the basic technical familiarity to functional neuroimaging. My previous involvement in the Optical Radiology Lab allowed me to observe day-to-day decision-making in creating a new functional neuroimaging technique. I have, in other words, first-hand access to the actual trajectory of technical development and to the thoughts and considerations that go into it. On the other hand, as someone who came from philosophy and had no prior exposure to functional neuroimaging, the learning experience as a complete outsider highlighted the practices and beliefs that would otherwise seem normal or intuitive to experienced practitioners. In short, I am a familiar stranger to functional neuroimaging. I enjoy both the privileged position of understanding the functional neuroimaging research community from the insider’s perspective and the curious eyes to question the community’s commitments and beliefs from the outsider’s perspective.

My third, structural, perspective is a perspective from the digital humanities. In this dissertation, I trace the history of an instrument for more than half a century with particular focus on the technical, social, and normative challenges that researchers went through. While I rely on narrative history to recount the efforts and challenges of a few research teams in the U.S., one always wonders if my choice of the research teams is reflective of the history of PET. In particular, I admit that my choice of those teams is partially shaped by convenience: my choice of the WUSTL team and the Massachusetts General Hospital (MGH) team is partially because I work on my dissertation at WUSTL and my family resides in Boston, close to MGH. To ensure that my history and my philosophical discussion is at least reflective and somewhat representative of the full history of
PET, I have to rely on more than just narrative history. After all, recreating and examining the comprehensive narrative history of more than a 100 research teams across the world and spans more than half a century is beyond my capacity. Therefore, to understand and analyze the history, especially how technical challenges interacted and shaped social and normative challenges (and vice versa), I rely on a bibliographic and computational approach. I have collected and analyzed the metadata, including the authorship and reference list, of more than a million articles both directly on PET and cited by articles on PET from the largest science journal article indexing service, Web of Science. The analysis reveals the structure of research and publication in PET that no individual researcher could process on her own. It offers a structural and large-scale perspective to the functional neuroimaging community, such as the diversity of academic disciplines that PET researchers tried to apply PET to, and how conquering specific technical challenges led to the next iteration of instruments across more than 50 years. The use of a bibliometric, computational, and structural approach broadens the historical coverage of my dissertation and hopefully ensures that my historical and philosophical analyses are more reflective the history of PET.

In short, with my structural, familiar, and foreign approaches to the historical episodes of PET, I am equipped to interrogate the history of PET and reveal insights into the development patterns and normative concerns in the development of PET and then the ensuing development of functional neuroimaging.

1.5 Chapter Summaries

Chapter 2 is an evolutionary chapter. It argues that an evolutionary metaphor is fitting and useful, but not literal, as a model of technology-driven scientific change. I argue that the development of PET and the technology-driven scientific change possess key elements of a process of evolution by natural selection. I also argue that, not only do multiple elements of an evolution by natural selection process have their correspondence in a technology-driven scientific change process, but also that those elements are best understood together theoretically as an evolutionary process. I
then suggest some further analogies that will become especially salient in the subsequent chapters: modularity, niche construction, and community ecology.

Chapter 3 looks at one of the core norms governing the evolution of PET: Signal-to-Noise Ratio. To do so, I focus on the development of the scintillator, a gamma-ray detection module. The history traces back to Hartmut Kallmann, a German physicist, and his research effort before and after the World War II in extensively testing out possible crystals to find those that could most effectively capture gamma rays and turn them into visible light for easy recording. I focus on how Kallmann pursued a higher signal-to-noise ratio for his gamma-detection device. I argue that concern for signal-to-noise ratio crosscuts and is more than interests in accuracy and reliability and deserves consideration as an epistemic virtue of detection technologies in its own right. Signal-to-noise ratio is a measurement of how well detection instruments can separate useful data from useless data. Increasing signal-to-noise ratio means detection instruments can better extract useful data from the general pool of information. This norm played a significant role in the early development of gamma ray detection primarily because gamma rays were invisible, hard to stop, and otherwise extremely hard to extract and separate from other forms of radiation, such as alpha and beta rays.

While Chapter 3 focuses on the norm of increasing signal-to-noise ratio, Chapter 4 explores the detection norm of increasing sensitivity. I argue that concern for sensitivity, like signal-to-noise ratio, goes beyond interests in accuracy and reliability and deserves separate consideration. The historical focus in Chapter 4 is the development of coincidence detection at MGH. Coincidence detection is a second key modular component of the PET technology overall. In 1950s, researchers at MGH realized that having a pair of gamma-ray capturing cameras, rather than just one, could significantly increase the sensitivity of their detection system. Sensitivity measures the extent to which a detection system extracts and utilizes useful information from the entire pool of useful information. In subsequent iterations of the technology, researchers at MGH further increased the number of cameras without going through fundamental redesigns with the goal of, again, increasing
sensitivity to the available signal. This module, coincidence detection, continued to play a role in the effort to locate the gamma ray signals from biological tissues, a key feature of subsequent neuroimaging technologies.

In Chapter 5, I use bibliometric analysis as evidence for a causal-historical account of how PET survived its immature and surprisingly unpopular years. I argue that the modularity of PET enabled PET to survive its immature years. Rogers (Rogers 2003) and other researchers (for example: Verganti 2009) on the dissemination of innovation have focused primarily on the phenomenon of the dissemination of innovation, especially after an innovation is proven and becomes more-or-less self-sustaining. Their understanding of the dissemination of innovation lacks a causal account of how new, immature, and unproven technologies may sustain themselves till they are proven and accepted. To provide such a causal-historical account, I use bibliographic analysis to illustrate how the PET field changed after research at MGH in the 1950s. The MGH team demonstrated that PET was a useful technology, but they failed to find a new use case for the technology, especially a use case that would justify a 4 million USD (in today’s currency) price tag. Analyzing the literature based on meta-information such as citations, author affiliations, etc. reveals that researchers around the world tried to modify the technology for uses for heart disease, lung disease, cerebral blood pressure, and for animal studies. Although each project was small, collectively there was enough interest to sustain the presence of PET in the medical research world. PET was able to find uses in such a diverse range of medical detection applications partially because of its modularity. Thanks to modularity, researchers could configure the modules for new use cases and compete under a wide range of data generation norms all at the same time. PET survived its immature years due to modularity.

In Chapter 6, I argue that another norm, increasing resolution, also warrants philosophical attention, because it demonstrates that information acquisition and empirical findings partially depend on the scale and frequency of sampling. In the historical section of this chapter, I recount the endeavors at
WUSTL that finally led to a successful new use for PET – using PET to detect cerebral physiological changes in real time. This development brought PET to the medical research mainstream. The chapter is focused on the choice among the competing governing norms in detection system design discussed in Chapters 3-5: should the research team purse better spatial resolution, better temporal resolution, better sensitivity, or maybe all at once? The chapter demonstrates that, as often occurs in evolution by natural selection, technological developments can shape their competitive landscape. Like a niche construction process, researchers at WUSTL created the condition for its own success by focusing on temporal resolution (“speed”).

In the final chapter, I summarize my main findings and discuss directions of future work growing out of this preliminary investigation.

1.6 Conclusion

In this dissertation, I pursue a historically grounded and philosophically motivated study of Positron Emission Tomography, through archival research, interviews, bibliometric analysis, and philosophical reflection.

I point out that the history of Positron Emission Tomography demonstrates at least three aspects, following three out of the four challenges mentioned by Hacking, through which scientific instrumentation is philosophically interesting: 1) as a technical challenge: scientific instrumentation is a distinct technical challenge; 2) as a normative challenge: scientific instrumentation follows its own engineering norms, such as signal-to-noise ratio, sensitivity, etc., often not covered by existing literature; 3) as a social challenge: the availability and specifics of tools often constrain and shape future theoretical research.

The goal of the dissertation is to start a new chapter in the recentering. I use the history to demonstrate that engineering and scientific instrumentation follow their distinct norms. And scientific instrumentation, not just the development of theories, can induce genuine scientific change. I
hope my historical and philosophical investigation can serve as a call for action for novel research programs to carry on Hacking’s proposal and further investigate how scientific instrumentation plays central roles in science.
Chapter 2

The Evolutionary Framework of Scientific Instrumentation

2.1 Introduction

In this dissertation, I focus on the norms that govern, guide, and inform scientific instrumentation and the specific mechanisms for the development of scientific instrumentation. As I have discussed the norms in the previous chapter, I dedicate this chapter to the mechanisms. While each topic takes up a separate chapter, they are by no means separate in themselves: norms guide the specific mechanisms for scientific instrumentation; scientific instrumentation informs and changes norms as well.

The historical and philosophical literature has no shortage of proposed mechanisms for the development of scientific instrumentation. However, most of the philosophical accounts assume that scientific instrumentation is primarily a top-down process: engineers come up with a general idea or goal, develop the intention into a detailed design, and then break down the design into concrete, manageable components before realizing their design piece by piece. This account of scientific
instrumentation has its merits, primarily in explaining where the functions of instruments come from: functions come from the engineers’ intentions and designs, realized in material objects.

But this account has a descriptive challenge: it is not a fitting description of a large variety of scientific instrumentation processes, including the history recounted in this dissertation. It is not that the development of Positron Emission Tomography (PET) does not involve human design or human intention. Any man-made device is imbued with human intention. But in many cases the development process was more “opportunistic” than “planned.” Researchers working on PET did not start with initial intentions or goals and realize them step by step. Instead, they identified research goals and problems based on technologies available to them. And they created technologies based on research goals and problems they intended to pursue. Throughout the iterative process, both the available technologies and the goals changed as research interests came and went, research teams formed and dispersed, and the world changed too. In this iterative process, a group of researchers might start with one project. Thirty years later, a separate group of researchers might pick up the same threads but end up in a place completely not imagined by the pioneers thirty years earlier. In my dissertation, I recount a history in which a researcher started the project by finding new materials to detect gamma rays. Half a century later, a team of researchers developed PET that could perform in vivo scanning of the brain, using technologies traceable to the earlier research half a century before. At every step, researchers injected their intentions into their scientific instruments and into the iterative process. But if we look at the history, we find that the developmental path of scientific instruments is no longer a reflection of any clear intention: the instruments have been modified, re-used, and exapted way too many times to have any recognizable grand intention that guides the evolution of the instrument over the course of its history.

In this chapter, I venture, instead of the “top-down,” “planned” account, another developmental mechanism of scientific instrumentation based on an evolutionary analogy. The use of an evolutionary analogy perhaps will raise some eyebrows, because of the long history of appropriating and
misappropriating evolutionary theory. Within a few decades of the publication of *On the Origin of Species*, adopting evolutionary theory to describe and explain social, instead of biological, processes was already a cottage industry. The cottage industry consisted of some of the less reputable “Social Darwinism” elements as well as serious efforts to find parallels between social and biological processes, such as efforts to find an evolutionary basis in politics (Bagehot 1872), in ethics (Alexander 1892), and in economics (Veblen 1898).

But unlike most of the previous attempts to broaden and generalize evolutionary theory from a biological theory to a social theory, I use evolutionary theory as a metaphor. Citing the distinction made by Hodgson, my intention in this dissertation is not to show that evolutionary theory, in a fairly literal, general, and abstract although modified, sense, can describe social processes. Instead, evolutionary theory offers a point of reference, as an analogy and a metaphor (Hodgson 2008). I argue that there are enough similarities between the structure and components of evolutionary theory and the developmental process of scientific instrumentation I describe. Due to the similarities, references to evolutionary theory offer a useful viewpoint and an overarching, organizing theme to organize, describe, and crystallize the process of “bottom-up,” “opportunistic” scientific instrumentation covered in this dissertation.

To make the case that evolutionary theory offers a useful analogy to organize the relevant scientific instrumentation history in this dissertation, I divide the chapter into two sections. First, I critically review the existing philosophical literature on design and scientific instrumentation. I argue that, while it is a useful and important philosophical account of scientific instrumentation, it is not an accurate description of the history of PET. I then propose the evolutionary analogy as a source for a philosophical account of bottom-up, opportunistic scientific instrumentation process.
2.2 The Design Account of Technology Development

The role of human intention and its consequences in making material objects have been a major focus in the philosophical discussion of technology development and engineering. In particular, a group of European philosophers have systematically developed their account of technology development based on intentional design: scientific instruments are material objects that realize human intention, through a design process (For example: Kroes and Meijers 2006; Houkes, Vermaas, and Ridder 2005; Vermaas et al. 2013). The group primarily takes their inspiration from Newell and Simon to propose their own goal-driven accounts of design in technology development. In this section, I describe the group’s account of scientific instrumentation, which I refer to as the “design account of technology development.” I argue that, while the account has its merits, it is empirically inadequate to describe common scientific instrumentation processes.

Newell and Simon argue that cognitive process of problem solving is primarily a search process from the initial state to the goal state (Newell and Simon 1972, p. 72). The goal state is the ultimate, intentional state of knowledge that one wants to achieve. The initial state is the state of knowledge that one begins with. The problem, as defined in Newell and Simon’s book *Human Problem Solving*, is a situation in which the person wants to arrive at the goal state but does not immediately know the series of actions to perform in order to arrive at the goal state (p. 72). The process to address the problem, then, is to navigate and search in a problem space – the solver’s representation of the task to move from the initial state to the goal state, within available means and under restrictions such as limited time, to eventually reach the goal state. In other words, the process to move from the initial state to the goal state is goal-driven and intention-driven because the goal state guides, drives, and is the ultimate aim of the search process.

A group of philosophers, many of them associated or have been associated with the Delft University of Technology (TU Delft), adopt the Newell and Simon framework to explain and describe the
developmental process of scientific instrumentation. In particular, I focus on Houkes, Kroes, and Vermaas. According to Kroes, the product of design is material artifacts and the design process is to realize a function with physical materials. Kroes notes that “designing may be described as opening up this black box and filling it with a physical structure that is able to perform the required function.” (Kroes 2012, p. 137) By the “black box,” Kroes refers to a coarse-grained description of function, such as “playing a CD” or “unscrewing the screws.” The intended function, in effect, is the goal state that the designer wants to realize. The design process is to arrange physical materials to realize the coarse-grained description of function, or the goal state, in real life. Hence says Kroes:

From the point of view of the object of design, an engineering-design process may therefore be taken to start with (a description of) its function, its intended behaviour, and end with (the description of) a physical structure that realises that intended behaviour (Kroes 2012).

Houkes and Vermaas disagree with Kroes on the product of design. They want to broaden Kroes’s concept of design. Houkes and Vermaas think that the product of design is a use plan – a series of actions to realize given goals (Houkes and Vermaas 2010, pp. 18-21). For example, maybe my goal is to “play a CD.” But I do not need to arrange physical materials, at least not literally, to create a CD player. A series of actions – such as going to an electronics store and purchasing a CD player, can realize my goal equally well. Houkes and Vermaas note that “designing is primarily – sometimes even exclusively – constructing and communicating use plans.” (p. 26) In other words, a use plan is what engineers are designing for. The creation of new material artifacts is not in general necessary as a part of designing.

Therefore, Houkes and Vermaas think that designing a product is a “subtype” of designing in general (p. 26). In specific cases, designing a material product could be a necessary means to the proposed use plan. But many use plans do not require designing new material products (pp. 29-30). One such example is to creatively use paperclips to pick the lock if I happen to lock myself out.
Despite the differences, Houkes and Vermaas agree with Kroes that design is a goal-driven activity. The following is the first few steps of Hourkes and Vermaas’s conceptualization of engineering design process. Houkes and Vermaas do not intend their steps to be an accurate description of the design process because the process often involves many feedback loops and other details left out of the picture. They do think that it captures the important conceptual steps in design (Houkes and Vermaas 2010, p. 29):

Goal contribution:

D.1. designer d wants to contribute to realising goal state g.

D.2. goal adjustment: d believes that g’ is the closest viable approximation of g.

D.3. d intends to contribute to realising g’.

Plan construction:

D.4. d intends to construct a new plan p for realising g’.

The construction of the use plan is conceptually in service of the goal. Although Houkes and Vermaas do not rule out that the use plan may shape the goal, ultimately the plan should contribute to fulfilling the goal.

According to Kroes, Houkes, and Vermaas, engineering design and technology development is primarily a goal-driven, top-down process. Like Newell and Simon’s account, to design and develop technology, the designer and engineer comes up with a goal state and then devises a plan to realize the goal.

Kroes, Houkes, and Vermaas favor this top-down account of engineering design partially because it helps to explain why instruments and technologies have functions. They move away from questions concerning functions in natural, biological processes (such as what the function of the heart is) to functions in artifacts and human made processes, such as what the function of a CD player is (for example: p. vi). Why is the function of a CD player to play CDs and why is it malfunctioning when it fails to play CDs? For Kroes, Houkes, and Vermaas, the short answer is that it comes from
designers and users’ intentions. It is because the designers put the materials together, dubbed the end product as a CD player, in order to play CDs (Houkes and Vermaas 2010, p. 88; Kroes 2012, p. 118). A CD Player is the material realization of a goal state to play CDs.

The challenge of the aforementioned philosophical view of engineering design, I argue, is that it does not adequately describe the complicated, often non-goal driven reality of the design process. I am referring to the lack of attention to “nonroutine” problems. I hereby resort to Visser’s extensive review of engineers and scholars of engineering’s discussion of “nonroutine” problems in the nineties (Visser 2006, pp. 135-136). Routine problems are primarily problems that have a well-understood sequence of actions to solve problems or at least a well-understood decision-making and computational method to do so. One such routine problem is to get an A in my calculus class. One routine engineering design problem is to DIY a new computer for myself.

But many, if not most, engineering design problems are nonroutine. They are nonroutine in the sense that they do not have well-defined goals. Nor do we have good decision-making or general strategies to arrive at those goals (pp. 135-136). One daily nonroutine problem is to get rich. The goal is vaguely defined – how rich are we talking about? And there is no good decision-making process or general strategies to get us there.

The design account of technology development fails to incorporate nonroutine problems. Given that nonroutine problems often do not have well-defined goals, engineers cannot carefully define and articulate the goal and come up with steps to achieve the intended goal. Given that nonroutine problems often do not have existing, proven decision-making or general strategies, it is difficult for engineers to even come up with steps, had the goals been clear and well-defined. The design account of technology development simply cannot adequately describe how people solve nonroutine problems.

The history of Positron Emission Tomography – the central case in my dissertation, largely involves trying to solve nonroutine problems. In the history of PET, the trouble for the design account
further compounds because researchers often radically change their goals and problems. Instead of searching strategically in a solution space to get from the initial state to the goal state, researchers are dealing with changing and often ill-defined goals. The history is not a top-down, goal-driven story.

The history starts with a simple, but ill-defined goal. After World War II, institutions such as the U.S. Atomic Energy Commission (AEC) and scientists worked in nuclear research faced the challenge of justifying the continued existence of large-scale nuclear research (Creager 2013, p. 1). They wanted to find exciting, valuable civilian uses of nuclear power. AEC funded the first cyclotron at Washington University in St. Louis and many ensuing nuclear research projects. The goal shared by AEC and many involved researchers was essentially to find something to justify continued, large-scale nuclear research. It was perhaps as vague as a goal can be. And there is no good general strategy to achieve the goal. But out of the vague goal and research agenda came the birth of Positron Emission Tomography and many other innovations.

In many episodes of the history of PET, similarly ill-defined goals have been the norm instead of the exception. For much of the 1950s-1970s, researchers at many research institutions were looking for academic uses for their new PET instruments. Their goal involved primarily learning how to find a suitable academic use for the new instrument to answer new academic questions and attract academic interest. Like the previous goal, this is not a clearly defined goal or that involves a clear, general strategy to realize this goal.

In other words, Positron Emission Tomography, along with the steam engine, the rocket, the microwave, and many other technological successes in our life, was not the product of goal-driven, routine problem solving. They represent a distinct category of technology development and scientific instrumentation.

I suggest that philosophers need a distinct account for technology development and scientific instrumentation that does not require clear overarching goals, strategies for problem solving, or
2.3 The Evolutionary Framework

This dissertation is a case study of the process of scientific instrumentation. Philosophers of science are famously more interested in the context of justification than in the context of discovery. I do not mean that the context of discovery does not receive attention by philosophers (one such example being Darden: Darden 1991). I do mean that the lingering Popperian influence persists in the background: the context of discovery is too irrational, too messy to properly belong to philosophy (Popper 2002, p. 8). Maybe psychology or the social sciences in general are more fitting homes for the context of discovery. At the same time, an a-rational and fairly “messy” process has received enormous attention by philosophers of science: the evolutionary process.

I argue that an analogy with evolution by natural selection can serve as a useful and illuminating overarching theme for understanding and organizing elements of PET’s developmental process and perhaps that of many other instruments and technologies. I first review the general evolutionary framework and point out many philosophers try to extend and generalize the evolutionary framework outside the narrow biological realm. I then argue that it is sufficient, at least for illuminating and organizing the mechanisms for scientific instrumentation, to use the evolutionary framework as a useful analogy. In particular, I emphasize the parallel between modularity, niche construction, and community ecology in evolution by natural selection and the process of scientific instrumentation. I hope that, in addition to a general model of scientific instrumentation, this dissertation is also able to demonstrate another source of philosophical value of the context of discovery: by its analogy to evolution by natural selection.

To get the analogy up and running, consider the abstract understanding of natural selection that is commonly accepted in the philosophical literature. Philosophers such as Darden and Cain (Darden
and Cain 1989), Griffiths (Griffiths 2006), Hull (Hull 1988b), Kitcher (Kitcher 1993), Skipper (Skipper 1999), and Sterelny (Sterelny 2012), among others, have all been trying to generalize and expand the evolutionary framework to understand and explain sociocultural and philosophical phenomena, such as the creation and transmission of ideas. They rely on roughly the same process: they abstract and generalize from the concentrate content of evolution by natural selection and argue that the same theoretical framework and insight can be instantiated in a different set of concrete content.

I am using two examples to demonstrate the shared approach among philosophers. Darden and Cain and Hull offer type-level abstractions of natural selection that can accurately describe various evolution by natural selection processes. Darden and Cain propose that their type level abstraction is intended as a schematic outline that identifies key components and processes at a higher level, to be further instantiated in clonal selection, neural selection, etc. (Darden and Cain 1989, p. 107). Hull similarly proposes an abstract evolutionary model that should describe biological, social, and conceptual developments (Hull 1988b, p. 13).

In addition to similar processes, Darden and Cain and Hull identify similar key components and processes in their type-level abstractions of the evolution by natural selection process. Darden and Cain’s abstraction consists of four key components and processes (Darden and Cain 1989, p. 110):

- Precondition of individuals with heritable variations and similarities and factors in the environment that can prefer an individual (or a group) over another
- Causal interaction between the individuals and those selective factors
- Differential interactions have differential results in the short term
- Short term results translate into long term differential results replicated and potentially propagated among the population

Hull similarly identifies four key components and processes (Hull 1988a, p. 134):
“a process in which the differential extinction and proliferation or interactors cause the differential perpetuation of the replicators that produced them.” (Italics original)

Here, in the quotation, Hull suggests that a type level abstraction of evolution by natural selection should include:

- A replicating population
- that can in the long run (“perpetuation”)
- benefit from or be harmed by (“extinction” and “proliferation”)
- causal interactions (“cause”) with other features in the world

In short, they all suggest the following components and processes as central to an evolutionary framework:

- Unit of Selection: a replicating population of varying units that have heritable features to be selected on
- Fitness Currency: there is a criterion or many criteria to benefit or harm the population
- Causal Interaction: there has to be causal interaction between the population and environmental factors
- Long Term Effect: the causal interaction should, through fitness, lead to long term differential results on units of selection

In this dissertation, I opt to use “evolutionary framework” to refer to their type-level abstraction of key elements to the evolution by natural selection process.

With the high-level abstraction in hand, Darden and Cain and Hull instantiate the abstraction to other phenomena. Darden and Cain instantiate the abstraction in other biological phenomena and processes, such as clonal selection and neuron group selection (Darden and Cain 1989, p. 125). They
demonstrate that it is possible to use the abstraction to render plausible explanations in fields outside natural selection and they suggest that their abstraction can apply outside biological processes too (Darden and Cain 1989, p. 127). Hull is more interested in instantiating the abstraction outside the narrowly defined biological realms, especially into social phenomena. He suggests that an evolution by natural selection process is an accurate description of the development of scientific ideas and theories (broadly construed) (Hull 1988a; 1988b, p. 9). Hull identifies the relevant key components and processes as follows:

- **Unit of Selection:** the replicating population of ideas (Hull 1988a, p. 141). Although ideas exist in scientists’ mind, they do exist for a long time and ideas give rise to ideas – or that they “mutate” and “reproduce” through scientists’ efforts.

- **Fitness Currency:** Ideas benefit or are harmed through the crediting process in science (p. 129). The worst thing we can do to ideas is to ignore them. And ideas “benefit” in the sense that they are cited and integrated into later ideas.

- **Causal Interaction:** Scientists interact among themselves and with the natural world to generate, improve on, and sometimes abandon ideas (p. 141).

- **Long-term Effect:** Ideas can be transmitted or forgotten in the long run (p. 140).

Hull’s application of evolution by natural selection on ideas runs into many theoretical challenges. The challenges primarily stem from Hull’s desire to label his model of ideas and concepts literally “evolutionary.” The most common challenge is that ideas are intentional products, but the evolutionary process is mindless. Reviewers point out that Hull ultimately does not respond to the issue of intentionality in his evolutionary model (for example: Oldroyd 1990). But Hull’s efforts to label the creation and transmission of ideas “evolutionary” break down at many other places too. For example, Rosenberg points out two major dissimilarities between the evolution by natural selection process and the process of creating and transmitting ideas (Rosenberg 1992). First, abstract entities
such as ideas are types, but natural selection takes place at the token level. Second, it is challenging to demonstrate how abstract entities have causes and effects.

The challenges with generalizing and instantiating theories of evolution by natural selection theory into other realms go beyond mere mismatches between features of the evolution by natural selection process and that of other processes intended to be modeled by the evolutionary framework. Another critical challenge to apply the evolutionary framework is on determining “essential” features of theories of evolution by natural selection. Is being “mindless” an essential feature of the evolutionary framework? Is some form of population generics an essential feature of the evolutionary framework? To use another example, Gould was famously critical of the early uses of evolutionary explanation in sociobiology. Gould thought there was a mistaken equivocation of evolution with adaptationism. As the result, the early versions of evolutionary explanation in sociobiology overly focused on an adaptationist explanation of phenomena while ignoring other mechanisms in evolutionary theory: not every trait, feature, or phenomenon was the product of adaptation (Gould and Lewontin 1979; Gould 1986).

I instead argue that the evolutionary framework is merely an extremely useful metaphor to understand the process of scientific instrumentation. I do not proclaim to understand the “essential” features of the evolutionary process. Nor is it needed for me to claim in this dissertation that the process of scientific instrumentation is another (potentially poor) instantiation of a generalized, abstract evolutionary framework. Instead, the similarity between the evolutionary framework and the scientific instrumentation process already offers enormous insight into how tools and instruments are developed. First of all, I argue that there is enough similarity between the natural selection process and the scientific instrumentation process for the analogy to work. Second, I admit that, at multiple points, the analogy breaks down and I do not literally claim that scientific instrumentation process is a natural selection process. Lastly, I point out the pragmatic value of invoking the evolutionary framework: contemporary theories of evolution by natural selection are not one theory per se, but
a hodgepodge collection of components. Many of the components do find their analogy in the scientific instrumentation process, although the latter is not strictly a process of evolution by natural selection. As I invoke more and more components to describe and explain how an opportunistic and bottom-up scientific instrumentation could work, it makes sense to use an evolutionary metaphor as the overarching theme and organizing concept for the scientific instrumentation process. The evolutionary metaphor is not literal, but it offers a new way, different from the design account of scientific instrumentation, to think of how scientific instruments come into being.

First, there are enough similarities and parallels between the evolutionary framework and elements in scientific instrumentation for the metaphor to at least stand. Actually, perhaps the scientific instrumentation process fits the major pillars of the evolutionary framework better than Hull’s account of the creation and transmission of ideas.

- **Unit of Selection:** Each instrument and group of instruments. Instruments “replicate” and “inherit” in the sense that researchers most often build instruments out of existing components and as improvements or deviations from other existing instruments. Instruments “mutate” because researchers gradually modify their designs.

- **Fitness Currency:** Instruments “benefit” or “are harmed” by adoption and replication. Like ideas, the worst treatment to any instrument is to ignore it and the best fate is to be widely used in a variety of research scenarios.

- **Causal Interaction:** Researchers use instruments to interact with the natural world and interactions among researchers could induce changes in instrument designs too.

- **Long-term Effect:** Modifications of instruments could lead to their long-term popularity or cause them to disappear from current usage.

Unlike abstract ideas, each individual instrument is a token that can be subject to selective forces. Second, physical, mechanical instruments are indeed causally efficacious. On the surface, the
process of scientific instrumentation bears a closer relationship to evolution by natural selection than the development of abstract ideas.

However, the process of scientific instrumentation is not literally a natural selection process for at least two main reasons. First, the development process of scientific instruments is an intentional process: human intentions are present and guiding the entire process. While the dissertation wants to focus on how human intentions are scaffolded by what was available in the development process and that the original intentions are commonly a far cry from the end results, there is no denying that human intentions are everywhere in the process of scientific instrumentation. Instruments do not assemble themselves. Second, the development process of scientific instruments is at times “Lamarckian.” Unlike a Darwinian process in which learned features and capabilities in an organism’s lifetime are not passed down to the offspring, features acquired in the iterations of development are passed down to the offspring, if human researchers find those useful.

For those reasons, I do not claim that the scientific instrumentation process is literally a process of evolution by natural selection. Instead, the evolutionary metaphor, using points of similarities, is already a powerful means to open up a different way of thinking about scientific instrumentation. In particular, in this section, I will discuss three themes of theories of evolution by natural selection that I think offer good insight into the scientific instrumentation process.

### 2.3.1 Modularity

Earlier in this chapter, I mentioned that there is no one evolutionary theory: research on evolution by natural selection consists of wide range of collections of component theories in the evolutionary process. Here, I am diving into one such component, the idea of modularity in evolution by natural selection. I introduce the idea and then point out that modularity is linked to evolvability. I then connect modularity in the natural, biological world, to the developmental process of scientific instrumentation. In particular, I argue that modularity allows researchers to quickly devise and
improve on their instruments without a deep understanding of the high-level theories of their target phenomena.

Modularity, in essence, is a feature of complex systems that it can be broken down into functional sub-systems and components. Following Simon’s work (Simon 1962, p. 468), contemporary biologists propose the following four characteristics for modularity (Callebaut and Rasskin-Gutman 2005, p. xv):

- Hierarchical Organization. Modularity takes place in a system that can be decomposed into sub-systems that span a hierarchy of parts and components.

- Near-Decomposability. “Modules” can be reasonably cleanly decomposed, i.e., isolated from one another through reasonably intelligible and limited interfaces and replaced with functional equivalents.

- Quasi-Functional Independence. “Modules” should be more or less self-contained functional units. It is understood that modules do need input and output from, and sometimes in other ways depend on other modules. Therefore, independence is a gradual degree and modules mostly have only quasi-, instead of full, independence from other modules or the system.

- Recurrence. Research in homology suggests that modules are often recurring (Müller and Newman 2003). Biologists can often identify similar functional units across a very wide spectrum of animals.

Modularity is fundamental to evolution by natural selection because modularity enhances evolvability. Evolvability here refers to the extent to which a population can produce better fitting, more adaptive offspring in any environment. The story of Tempus and Hora, as retold by Simon, offers a clear analogy (Simon 1996, p. 188): Tempus and Hora are both fantastic watchmakers, but with a major difference. Tempus assembles holistically: whether something falls apart, Tempus has to start from scratch. Hora assembles piecemeal and modularly: if something falls apart, Hora only needs to
rework on the broken parts. Because Hora’s process is modular, Hora is able to produce more functional watches than Tempus. To put the story of Tempus and Hora in more theoretical terms, I am adopting a simplified version of John H. Holland’s theory (Holland 1992) to explain the relationship between modularity and evolvability. Assume, for a moment, evolution is nothing more than a search for better solutions in a massive solution space. Some fundamental properties of an evolving population should include: (1) it should be able to store the traversed path and lessons learned along the search; (2) it should be able to access the learned knowledge; (3) it should be able to exploit the knowledge while evolving for better solutions. It quickly becomes clear that a modular system would fare better in such a search task: it can retain the working modules (hence storing and using what works) and switch out the harmful modules and test out new modules (hence searching for better solutions). Replacing the bad modules does not or does minimal harm to the other working modules. In comparison, a non-modular system would need to throw away what works and go through wholesale re-design if it wants to get rid of what does not work.

The evolvability point seems to apply perfectly well in the development of scientific instruments, especially through the case of PET. First of all, while developing scientific instruments is not simply a search in a massive solution, the process has some similarity to the natural selection process. Instruments can “benefit” or be “harmed” insofar as they are used or ignored by other researchers. Researchers developing instruments want to acquire more recognition, credit, and adoption of their instruments. But figuring out how to accomplish that is a major challenge. As I am going to describe in Chapter 5, researchers working on PET were frustrated by the lukewarm reception: merely a handful of publications were using their PET instrument. They wanted to find a use case for PET to attract widespread academic interest and they finally achieved the goal by focusing PET on real time cerebral changes in vivo.

Second, instruments are literally modular. For most pieces of scientific instruments we know, we can retain the working modules and replace the non-working modules. As I am going to describe
in Chapter 3 and later chapters, PET’s design is completely modular. Its core components involve a crystal to turn invisible radiation into visible light, a photomultiplier to amplify the minuscule amount of light, a camera to capture the light and convert into electric signals, and then a computer system to analyze and process the signals. Of course, PET has to be paired with specific radioactive isotopes to detect their radiation. Researchers can switch in the out the crystal to improve on the resolution of the system and the speed of each scan. Researchers can change the photomultiplier and camera to improve the signal-to-noise ration of the system. Researchers can also change the choice of isotopes to study many aspects of human, animal, and plant physiology.

In the dissertation, I argue that modularity is a key success factor for PET. As I alluded to in the introduction and will expand in later chapters, because of modularity, researchers were able to swiftly adopt new developments in nuclear research into existing medical instrument designs. They were able to re-use crystals and detectors designed to count gamma rays for nuclear research into the same core parts of studying human biology. Also because of modularity, researchers were able to come up with instrument designs that may potentially work in a wide variety of medical disciplines and quickly identify potential venues for adoption, such as all the way from studying small animals, studying human cardiovascular problems, studying cancer, and studying cognitive processes all with similar core technology. Modularity allowed PET to quickly pivot, instead of re-designing from the ground up, and finally find widespread academic interests despite decades of failure to attract sufficient interests. Lastly, modularity allowed researchers to pivot into fields without high-level understanding of the target phenomena. When researchers pivot into studying cerebral physiological changes, high level theoretical understanding of the physiological processes was still limited in the 1970s. But the lack of high-level understanding of the target phenomena did not prevent researchers from developing their instruments, because most key modules were already tested and verified in other uses and can be re-used for the new target phenomena.
2.3.2 Niche Construction

In addition to modularity, I introduce a second extension of the evolutionary metaphor, the role of niche construction theory in understanding the process of scientific instrumentation.

Niche construction theory is a theory built against Classical Darwinism that organisms are mostly passive participants in the evolutionary process – organisms merely respond to and are selected by evolutionary forces. Lewontin summarizes this idea in Classical Darwinism as follows (Lewontin 1982, p. 159):

Environment begins as alienated from the organism, which must then bring itself into conformity with the given world. This view of environment as causally prior to, and ontologically independent of, organisms is the surfacing in evolutionary theory of the underlying Cartesian structure of our worldview. In fact, it is impossible to describe an environment except by reference to organisms that interact with it and define it.

Organism and environment are dialectically related (p. 160)

In other words, in Classical Darwinism, the environment is ontologically independent from and prior to the organisms being selected. Organisms merely react to the environment they are in. Organisms evolve or go extinct depending on how well they respond to the selective pressure in the environment.

Roughly around 1960s, biologists came together to develop an alternative to the Classical Darwinian view that the organisms are merely passively selected on. The earliest formal statement I can find was published in 1959, when Waddington pointed out that animals often faced “a much wider range of environmental conditions than they are willing to inhabit. They live in a highly heterogeneous ‘ambiance’, from which they themselves select the particular habitat in which their life will be passed. Thus, the animal by its behaviour contributes in a most important way to determining the nature and intensity of the selective pressures which will be exerted on it.” (Conrad Hall Waddington
In 1969, Waddington better formulated and expanded his thesis. He claimed that “the environment which exerts selection on one organism is influenced by the presence of other organisms; and as the other organisms change in evolution, so the environment of the first organism is altered, and it must evolve too.” (C. H. Waddington 2008, p. 262) In the 80s, following Lewontin and Waddington, Odling-Smee became one of the first biologists to propose a systematic account of niche construction. The niche, according to Odling-smee et al., is “the sum of all the natural selection pressure to which the population is exposed” (Odling-Smee, Laland, and Feldman 1996, p. 40). Niche construction, then, is when an organism modifies the relationship between parts of the organism and parts of the environment by actively changing parts of the environment, by perturbing the environment, or by relocating to a more favorable environment (p. 41).

Niche construction theorists point out that organisms actively choose or reconfigure their environment to their own benefit. Previously, I simplified the evolutionary process as a search in a massive solution space. According to niche construction theory, organisms do not always search for answers for existing questions. They can rewrite the questions so they, and to the disadvantage of others, already have the answers. They not only benefit from changing the environment in their favor, but also creating selective pressures for both their own species and their competitors, because the changed environment can last. For example, beavers build dams to create ponds and to protect themselves from predators. With the dam, now the environment potentially favors the beavers and disfavors the predators. For another example, birds probably first evolved feathers for thermoregulation. But once feathers were exaptated for flights and changed the competitive landscape, other body parts (such as chest muscle) went through secondary evolution to better facilitate flights (Gould and Vrba 1982, p. 11).

Niche construction theory is useful in understanding scientific instrumentation for at least two reasons. First of all, it emphasizes researchers’ role in actively reshaping the competitive landscape into which their instruments were received. Instead of merely appealing to other potential users of a
new instrument, researchers can change other users’ preferences or tap into underserved populations and find new niches to exploit for the adoption for their new tool. Second, it emphasizes how researchers can change how other, potentially competing instruments, will be received and evaluated in the future.

In the case of PET, researchers at WUSTL found an untapped environment. Being able to witness cerebral physiological changes allowed researchers to understand how regional cerebral physiological changes corresponded to human cognition in an entirely new way. Previously, lesion studies were the dominant method to understand how human cognition was realized in cerebral physiology. But they required lesions that researchers could not precisely control. And while lesions helped to establish the relationship between lesions and cognitive functions, they did not allow for real time observation of how other regions interacted with each other or were involved in daily cognition. The real time window to cerebral observation was hence a completely new page. Researchers at WUSTL quickly occupied the niche. And their success shaped the landscape so that multiple imaging techniques started to compete more seriously on the speed from then on. For sure, improving speed has always been desirable in instrument design. But improving speed was relatively a low priority among other norms and desiderata. It is because researchers at the time tended to think improving the resolution of images from instruments, instead of producing those images faster, could solve and generate more academic questions. For this reason, the norm of increasing speed was relatively of a lower priority. However, once the team at WUSTL demonstrated the sheer number and possibility of new questions to be answered by focusing on speedy scans, the norm of increasing speed became an important norm in neuroscience and guided the future generations of imaging instruments.

2.3.3 Community Ecology

In the previous subsection, I primarily focused on how a single community of organisms navigates the competitive pressure of its environment. In this subsection, I draw another evolutionary metaphor to understand how multiple communities interact with each other in a competitive environment. I
am referring to the discussion of community ecology. Community ecology is a broad collection of theories in biology. In this subsection, I only focus on the relationship between community structure and competition for resources and processes.

Researchers have long noticed that patterns of competition and organisms’ ensuing response shape the structure of ecological communities (for a few examples, see Tilman 1982; Chesson 2000; Pulliam 2000). But studying the relationship between competition and community structure has been difficult. Researchers need to track the static relationship, by identifying a myriad of relevant features, states, and processes of the environment and organisms. They also need to track the dynamic feedback relationship as the environment and organisms change in response to each other (Pulliam 2000, p. 351).

To process the complex, dynamic relationship between competition and community, researchers often rely on advanced mathematical modeling. Tilman is one leading example because of his systematic use of graphical and topological theories. He extends a long history of research to blend the study of ecological communities with graphical and topological theories, which he argues would provide the right level of abstraction and reveal patterns and assumption that otherwise may not be visible (Tilman 1982, p. 5). His tries to address a core question in biology: why are there so many species on earth (p. 3)? Especially why are there so many species that co-exist in an environment all competing for a very limited set of resources (p. 8)? The short answer is that those species are able to coexist in an equilibrium of resource consumption and highly specific community structure (p. 5). The identification of the equilibrium, then, comes down to intensive reliance on graphical and topological theories. The goal of Tilman’s research is to study, model, and predict the equilibrium, primarily based on his research on plants, with only speculative application on animals (p. 6).

In the models developed by Tilman and other researchers, they commonly share overlapping lists of features and states of the community and organisms. While a distinct and often unique set of features
and states determine the relationship between specific pairs of the community and competition, many features are recurring across communities and organisms. Some critical features include:

- **Resource**: substances or factor consumed by the organisms and can modify their growth rates (Tilman 1982, p. 11). Researchers often collect a number of features regarding one resource, such as its availability in the environment, its supply rate, and its interaction with other resources. One example of resource is sunlight for a large number of organisms.

- **Process**: environmental and organism interactions that are not consumed by organisms but also modify their growth rates (Chase and Leibold 2003, p. 14). Common examples of processes involve predation and nitrogen fixation.

- **Growth rate**: per capita rate of reproduction (Tilman 1982, p. 11). Researchers also commonly break down growth rate into birth rate and mortality rate.

- **Impact**: per capita rate of impact, by the organisms, on the requisite resource or process (Chase and Leibold 2003, p. 20). For example, cats have a major impact on the environmental availability of mice.

Similar to my earlier observation, the competition in scientific instrumentation has a striking similarity to the competition for resources. Scientific instrumentation research, for example, all require the funding and attention, which in a metaphorical way those research programs “consume.” Hence, we can draw a list of comparable features in scientific instrumentation:

- **Resource**: substance or factor consumed by research programs that modify their publication rate. Researchers commonly need funding, personnel, and attention, among others to perform research and publish.

- **Process**: larger industry trends that modify research programs’ publication rate. For example, the growth of cognitive neuroscience as a recognized subfield, associated with more journals, conferences, etc. enables a lot more publication in the field.
• Growth rate: publication rate of research programs.

• Impact: the impact of research programs on the general research community.

Connecting competition with community structure and using graphical and topological theories are both critical to the understanding of the development of PET. First of all, PET has long been in direct competition against other technologies at the time, such as X-ray CT. The competition caused PET to pivot to its own research and application niches. Second, understanding how exactly the competition shaped the current community requires more than a narrative history. PET is a collaborative effort. It takes more than just a few teams to bring the technology to the forefront of neuroscientific research. Due to limits of this piece of research, I could only compile the narrative history of primarily three PET-related research teams. Understanding the history takes more than just the history of those three teams. Similarly, understanding the relationship between competition and community structure often requires population level information and introduce the study of graphical and topological theories to understand how populations interact with each other. Therefore, in the last chapter in this dissertation, I rely on bibliographic information – “meta” information on the abstract, authorship, institutional affiliation, and reference data of publications to gain insight into “population level” changes in the history of PET. I intend to use the bibliographic information and insights from Tilman, Chase, and other working on community ecology to better understand how competition in scientific instrumentation shapes the structure of biomedical research community.

2.4 Conclusion

In this chapter, I argued for two points. First, I argued that the diverse range of scientific instrumentation process cannot be covered by the design account of engineering. Instead, philosophers need a new account to describe and understand the “opportunistic” and “bottom-up” process of scientific instrumentation. The design account of scientific instrumentation is a descriptive and normative account of the process of scientific instrumentation and the assignment of functions to
human artifacts. The design account is primarily popularized by Houkes, Kroes, and Vermaas, and their associates. The design account takes scientific instrumentation as a top-down, goal-driven, and strategy-driven process. Researchers and engineers normally start with a fairly well-developed goal of their design. They then break down the overarching goal into manageable components and gradually bring about the intended goal by manipulating the material world.

But the design account is empirically inadequate for a less goal-driven, strategy-driven process. Engineers often have unclear goals or have goals without any general strategy to achieve. As in the history of PET, engineers do not follow a process similar to the design account to develop new scientific instruments. Instead, they pursue a bottom-up, opportunistic process that requires a separate philosophical discussion of the process and normative considerations in the process.

Second, I argued the evolutionary framework offers a useful metaphor for such an “opportunistic” and “bottom-up” process of scientific instrumentation. I did not argue that engineering design or scientific instrumentation is an evolutionary process. I argued that the evolutionary framework was 1) similar enough to the bottom-up process of scientific instrumentation to draw parallels; 2) offers critical and applicable insights into the empirical description and normative considerations of the bottom-up process of scientific instrumentation.

I argued for the similarity between evolutionary framework and the bottom-up process based on some critical features of the process of evolution by natural selection. I focused on considerations of unit of selection, fitness currency, causal interaction, and long-term effect and argued that, more than superficially, the bottom-up process of scientific instrumentation has elements similar to the process of evolution by natural selection.

After justifying drawing the parallel between the process of evolution by natural selection and the bottom-up process of scientific instrumentation, I foreshadowed some key elements of the evolutionary process that I would explore in the ensuing chapters.
First, I emphasized the many evolutionary theories suggest that the modular structure of organisms likely benefits their evolvability. Modularity describes the features of organisms and instruments that they can be broken down into semi-independent functional components. Populations of modular organisms can evolve better-fitting components while preserving other components that are proven to be beneficial in a given environment. I extended the metaphor to scientific instrumentation and argued that the modularity was critical in the bottom-up process of scientific instrumentation. While the entire history of PET is an example of how modularity played an indispensable role in scientific instrumentation, in the ensuing chapters, I will focus on how modular modification was critical in the early decades of PET, especially in enabling the technology now known as PET to come into its shape from a diverse range of research and pursuits.

Second, I discussed the evolutionary theories of niche construction. Niche construction emphasizes the role of organisms in selecting and modifying their surrounding environment and competitive landscape. Organisms, instead of merely being selected on, can shape the selective forces and render the environment more favorable to them. In the history of PET, I relate many episodes in which researchers developed their instruments to shape and re-define the competitive environment. For example, in the 1970s, once researchers at WUSTL realized that PET would not compete under the current environment head-to-head against X-ray, SPECT, and other imaging devices, they pivoted and created a new research direction of imaging normal biological processes – especially cognitive processes, in vivo. Imaging cognitive processes in vivo, at the time, had little competition, and PET was able to answer many new and unique questions. In the favorable environment, PET received a lot more research interest, academic adoption, and usage in publications, despite its million-dollar price tag.

Third, I dug deeper into the relationship between niche construction and community ecology. Many theories of evolution by natural selection have pointed out that competition is a major factor in shaping the structure of ecological communities. For example, if two (or more) species compete
for the same resources in the same environment, and one species is more efficient with all the resources than the other species, then it is unlikely that the two species can co-exist. The less competitive species would either move to a different niche or evolve to consume the resources in the current environment differently. The connection between niche construction and community ecology matters for the history of PET because PET was not developed by one team or one research center. PET was not even “one” technology per se – different research teams created different variants of PET, although by-and-large those instruments shared the same or similar key principles. Because of the diversity in the community of PET researchers and instruments, following merely a few teams – teams at MGH or at WUSTL for example, provides a very incomplete picture of the community of PET and how it interacted with other communities of medical imaging technologies. Understanding how the PET community formed, developed, changed, and more or less converged on the study of cerebral and cognitive processes requires more than the narrative understanding of merely a few teams. Hence, I borrow from community ecology and rely on the bibliographic data on publications using PET to understand how the topology of the community PET changed in the 40 years between 1970s to the early 2000s.

In this chapter, I discussed the usefulness of an evolutionary metaphor in understanding the history of PET. The discussion is preliminary – to draw out the important similarities between the evolutionary framework and scientific instruments and to gesture towards the skeleton of applying an evolutionary metaphor to study the history of PET. I do intend to call attention to features of scientific instrumentation that would otherwise be neglected or invisible to us. Attention to those features provides an alternative schema that supplements the Design Account of Technology Development. In the ensuing chapters, I will recount the history of PET and embed the evolutionary metaphor into my analysis of the history. With substantial historical content, I hope I can make a convincing case that the evolutionary metaphor is useful and illuminating.
Chapter 3

1930s-1950s, The Invention of the Gamma-Ray Detector and the Epistemic Value of Signal-to-Noise Ratio

3.1 Introduction

In this chapter, I retell the history of Positron Emission Tomography from its beginnings in the 1930s. The “beginning” of PET is perhaps a misnomer because we can always find an earlier precedent that later generations draw inspirations from. I pick the beginning point in the 1930s primarily because this is when a key component of PET – its scintillator – was invented. Since the 1930s, the core principles of the scintillator have stayed largely the same.

In retelling the history, I also invite my readers to think about the normative values that guided the development of the scintillator. In the Introduction Chapter, I note that my philosophical goal is to understand the leading norms that guide, measure, evaluate, and improve the development of scientific instruments. Are those norms the same as well-known epistemic norms that philosophers
are already discussing? I hope the history can give a vivid example that the answer is a qualified no. For sure, researchers working on scientific instrumentation care about common epistemic norms that concern philosophers: accuracy, reliability, etc. (for a discussion of key scientific epistemic norms and their relationship with scientific realism, please see: Godfrey-Smith 2003, pp. 175-176). But instrumentation faces its unique challenges and has its own distinct normative values.

In particular, in this chapter, I focus on signal-to-noise ratio and its associated norm of increasing signal-to-noise ratio as critical components in designing and improving detection instruments. Signal-to-noise ratio, and the norm of increasing signal-to-noise ratio are critical to addressing a central and fundamental challenge in detection and in inference from data to phenomena: our world is too noisy and it is too hard to extract and recover useful information from our data. Signal-to-noise ratio measures the extent to which researchers can recover signal (useful information) from noise (useless, inferring information) from data generated by our detection technology.

The challenge of recovering useful information from data exists because no detection instrument can causally interact with and only with our target phenomena in the most ideal way. Undesirable causal interactions often create interfering data points that complicate our inference back to the target phenomena.

The quick brown fox jumped over the lazy dog.

Figure 3.1: Increasing signal-to-noise ratios

In Figure 3.1, both sentences have the same truth value. If there is a quick brown fox jumping over the lazy dog, then both sentences are true. But the first sentence is very easy to read: black, big,

1. I want to acknowledge Professor Hazlett for providing the example. Thank you!

62
well-spaced letters on a white background. The first sentence “stands” out from the background and we have no problem processing the meaning of and the information carried by the sentence. The second sentence is a struggle for many: small and narrow letters that are almost indistinguishable from the gray background. It is extremely difficult to process any information – or even to tell that the picture contains an English sentence. In both sentences, the relevant information is there and they have the same truth value. But it is hard to pick useful information out from and make the truth value judgment on the second sentence. The first sentence, in other words, has high signal-to-noise. The low sentence has low signal-to-noise. To better reconstruct the useful information researchers want to improve the signal-to-noise ratio of their detection instruments.

In short, signal-to-noise ratio is a measurement of recoverability of signal out of our detection instruments. The norm of increasing signal-to-noise ratio is then an engineering and epistemic norm to increase the recoverability of the signal from our detection technology. As I will dig deeper into later in the chapter, because signal-to-noise ratio measures the recoverability of useful input, it is indirectly a measurement of the quality of inference from existing data to the input – target, original phenomena of interest to researchers. Further, given the mathematical and causal nature in constructing the measurement of signal-to-noise ratio, the ratio allows researchers to identify causal improvements to their devices to increase the ratio and enable better recovery of useful information in detection instruments.

In order to tell the story of how signal-to-noise ratio played an important role in the instrumentation process of PET, I divide the chapter into two sections. The first section is historically oriented. I recount the early history of the gamma ray detector in the 1930s and the 1950s, which became a core module in Positron Emission Tomography. One significant aspect of the early history of the gamma ray detector was the pursuit of higher signal-to-noise ratio. The second section is philosophically oriented. I explain mathematically what signal-to-noise ratio means and discuss the nature and epistemic significance of increasing signal-to-noise ratio.
3.2 The History of Detecting Gamma Rays

This history section consists of two directions of research between the 1930s and 1950s, both research directions illustrate the fundamental importance of increasing the signal-to-noise ratio of the young gamma ray detector. In the first example, researchers throughout the 1930s, 1940s, and 1950s were looking for better materials to detect gamma rays. One main concern they had was to separate their recordings of gamma rays (signal) from recordings of other forms of radiation and visible light (noise). In the second example, researchers in the 1950s started to find ways to use gamma rays in medical diagnostics. One main challenge they faced was that issues and the skull would scatter gamma rays on their way out. As a result, it was difficult to infer where the scattered gamma rays originated. In short, they had to separate unscattered gamma ray readings (signal) from scattered gamma ray readings (noise). In each case, we see researchers struggling to filter out noise while leaving as much of the signal intact as they could.

3.2.1 Detecting Gamma Rays

In 1878, when Max Planck consulted his professor, Philipp von Jolly, at University of Munich for a career in physics, Philipp von Jolly thought it was a bad idea: he said the structure of theoretical physics was already completed with the discovery of the principles of thermodynamics (Heilbron 2000, p. 10). von Jolly could not have been more wrong. Kragh, when contracted by the Princeton University Press to write a history of physics in the twentieth century, quick learned that writing a comprehensive history of physics in the twentieth century in one volume is “not possible.” (Kragh 2002, p. xi) In fact, merely the first decade of the twentieth century was a watershed moment for physics. For example, Planck discovered the quantum of action in 1900 (p. 3). Einstein discovered the special theory of relativity in 1905 (p. 3). “This 20th century in physics began with a rush of new insights,” remarked D. Allan Bromley, renowned physicist and former president of the American
Among the fundamental new insights into physics around the turn of the twentieth century was the discovery of subatomic particles, or the starting point of nuclear physics as a subfield of physics. The very first subatomic particle, the electron, was suggested right around the turn of the century: J. J. Thomson, a physicist who was appointed Cavendish Professor of Physics at Cambridge University at the young age of 28, first theoretically described the electron in October 1897 (Kragh 2002, p. 44). Following, although eventually rejecting Thomson’s subatomic model, Rutherford identified alpha, beta, and gamma rays (they are particles as well, because of wave-particle duality) (Rutherford 1899, 1903).

Among the three subatomic rays – alpha, beta, and gamma, the gamma ray has a particularly interesting history. Röntgen accidentally discovered a highly penetrative form of radiation in 1896, which he named “X-ray” (Röntgen 1896). He quickly found medical uses for X-ray, creating a major industry of medical imaging. Rutherford and other physicists identified gamma rays by the turn of the century (Rutherford 1903). But it took researchers another decade, with many experiments and theoretical developments, to figure out that X-ray and gamma rays were the same thing (Bragg 1910; Friedrich, Knipping, and Laue 1912; Laue 1912; Rutherford and Andrade 1914). In other words, by the time theories and experiments on gamma rays were mature, the X-ray had already been used widely in British hospitals for more than ten years. On the one hand, it was a newly discovered and understood element of the subatomic world. It was a shiny new “toy” for interested nuclear physicists to play with and investigate. On the other hand, it was also a mature medical imaging tool, attracting the interests and attention of medical professionals and patients alike. Research into gamma rays was hot.

But quantitative measurement of gamma rays had always been challenging. Gamma rays are very energized and very penetrative. It was therefore difficult to find materials that could stop gamma
rays and turn them into easily countable forms of energy, which was what a gamma ray detector had to do (Hofstadter and Stein 1975, p. 13). Medical professionals had, before World War I, long been using films to stop and capture gamma rays. But such measurements were inaccurate and noisy: medical professionals could roughly see shades of gray as an analog of gamma ray intensity, but could not precisely count the number of gamma rays or fully separate them from the background noise.

In the 1930s, when the story of this section began, researchers found a promising new direction for the detection of gamma rays. They were aware of a wide range of chemicals that, when energized, generated visible light. Those materials could, in other words, convert invisible radiation to visible light. Because those chemicals “flash” when energized, they become known as scintillators.

The initial batch of scintillators were not particularly helpful in detecting radiation. Scintillators can convert radiation into visible light. But known scintillators in the 1930s were inorganic phosphors such as Zinc Sulfide, which exists as powder under normal circumstances. One single minuscule fleck of powder does not have the surface area to capture radiation coming from all directions. The pile of powder has too many gaps, too much air, etc. between flecks for good detection. Researchers hence embarked on a journey to identify better scintillators, preferably organic scintillators, because it was and still is easier to grow larger organic crystals.

The discovery of the sought-after organic scintillators is largely due to Kallmann’s contribution (Spruch 2008). Kallmann finished his PhD under Max Planck between the wars. He had enormous interest in basic physics, especially nuclear physics. But he was quickly out of work when Hitler came to power. Kallmann had Jewish lineage. Fortunately, because his wife was not a Jew and their kids were considered Aryan (he was only partially Jewish and his wife was designated “fully Aryan”), and because Aryan kids must have their both biological parents, Kallmann did not suffer much physical abuse. He was nevertheless unable to continue his research and made do with rather rudimentary neutron photography research in a former horse stable with a private fellowship from I.
G. Farben Company (Hine 1977, p. 868). Given that he was working on neutron photography, he had ultraviolet lamps in the stable. One day, when he entered the horse stable with the ultraviolet lamp on, he noticed that the seams of the accompanying technician's clothing were glowing. After calling the clothing manufacturer, Kallmann learned that the glowing content was Coronene. He immediately realized that organic solid scintillator would be a possibility (Oster 1966, p. 51).

But Kallmann had to wait until the end of World War II before he put his idea to experiments. He had extensive nuclear research experience and was worried that his discovery of an effective radiation detector might contribute to Hitler’s nuclear efforts. He kept his discovery a secret until Hitler was gone and then started a series of experiments (Hine 1977, p. 868).

In 1947, based on his earlier experiments, Kallmann invented the organic radiation detector. The detector primarily consisted and still consists of three main components. It has a scintillator crystal – the material of concern in this chapter, which flashes when energized by incoming radiation. Because the flash is very weak, researchers need a way to amplify the flash. Kallmann resorted to a mature, generally available module called the photomultiplier – which is still used in contemporary setups. In the tube-shaped photomultiplier, when incoming radiation hits its surface materials, those materials release a lot of electrons. The emitted electrons then go on to hit another layer of materials and cause them to release even more electrons. The process repeats inside the tube to amplify the scale of emission, metaphorically similar to an avalanche. Eventually, researchers rely on a camera not too different from our digital camera’s technologies to capture the amplified flash.

The challenge for a good organic radiation detector is that the energy from incoming gamma rays is not the only source of energy in the detection instrument. For example, the flash is visible light. But there are many other sources of visible light in the world. It is not possible to maintain absolute darkness in the tube other than the flash. So the photomultiplier will amplify other sources of visible light as well. For another example, other forms of radiation exist in the tube (such as universal background radiation). The photomultiplier will pick up those forms of radiation as well. For one
more example, both the photomultiplier and the camera are electronics and they have their own electrical currents. All of those forms of energy and more are picked up by the photomultiplier and then the camera, making it difficult to assess the actual amount of incoming gamma ray radiation. In short, the measurement coming out of the detection apparatus is a measurement of a mixture of all kinds of energy.

In order to separate the signals created by incoming gamma rays from data points created by other irrelevant radiation and currents, Kallmann and other researchers wanted to increase the signal-to-noise ratio of their detection instrument. While I will explain the concept of signal-to-noise ratio later in this chapter, I use an example to quickly demonstrate why signal-to-noise ratio is critical.

Figure 3.2 is an example from an online imaging textbook\(^2\). It is a graph of the readout of a detection instrument. In the readout, we see a persistent baseline noise. This is the noise generated by all kinds of irrelevant, intervening light, electrical currents, etc. If the sudden flash generated by incoming gamma rays is small (such as the “weak” pulse in the graph), then it would be hard for researchers to

\(^2\) http://www.optique-ingenieur.org/en/courses/OPI_ang_M05_C05/co/Contenu_12.html
tell that this peak indicated incoming gamma rays, instead of the natural variation of the background, baseline noise. If the sudden flash generated by incoming gamma rays is big (such as the “strong” pulse in the graph), then it would be clear that this is caused primarily by incoming gamma rays.

In order to create a “strong” pulse in the detection instrument, before and after his invention of the organic radiation detector, Kallmann and his colleagues experimented with a long list of potential crystal materials. In those experiments, Kallmann and his colleagues sourced a variety of inorganic and organic solids, such as $Zn_2SiO_4$, $ZnS$, $CaWO_4$, and Naphthalene, the effective content of moth balls structurally similar to Coronene and something that was easy to source right after the war. He and his colleagues then hit the solids with alpha, beta, and gamma rays and observed how much light they produced (Broser and Kallmann 1947; Immanuel Broser and Hartmut Kallmann 1947). Figure 3.3 is a summary of experiments as of 1949 (Kallmann 1949).

<table>
<thead>
<tr>
<th>Phosphor</th>
<th>Wave-length $\lambda$ (A)</th>
<th>Physical (relative to ZnS) $x$-ray $\lambda$ (A)</th>
<th>Technical (relative to ZnS) $x$-ray $\lambda$ (A)</th>
<th>Energy yield $\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha$</td>
<td>$\beta$</td>
<td>$\gamma$</td>
<td>1.93</td>
</tr>
<tr>
<td>ZnS-Ag</td>
<td>4500</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>ZnS-Cu</td>
<td>5200</td>
<td>140</td>
<td>160</td>
<td>160</td>
</tr>
<tr>
<td>ZnS/CdS-Cu</td>
<td>5900</td>
<td>46</td>
<td>46</td>
<td>46</td>
</tr>
<tr>
<td>CdS-Ag</td>
<td>7600</td>
<td>78</td>
<td>~80</td>
<td>~80</td>
</tr>
<tr>
<td>Zn$_2$SiO$_4$.Mn</td>
<td>5250</td>
<td>13.5</td>
<td>81</td>
<td>80</td>
</tr>
<tr>
<td>ZnO</td>
<td>5500</td>
<td>7.0</td>
<td>7.0</td>
<td>7.0</td>
</tr>
<tr>
<td>CaWO$_4$</td>
<td>4300</td>
<td>6.3</td>
<td>49</td>
<td>60</td>
</tr>
<tr>
<td>MgWO$_4$</td>
<td>4900</td>
<td>6.4</td>
<td>6.4</td>
<td>6.4</td>
</tr>
<tr>
<td>KBr-Tl</td>
<td>3600</td>
<td>6</td>
<td>50.4</td>
<td>50.4</td>
</tr>
<tr>
<td>Naphthalene</td>
<td>3600</td>
<td>20</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td>Diphenyl</td>
<td>3600</td>
<td>6.0</td>
<td>52.5</td>
<td>52.5</td>
</tr>
<tr>
<td>Phenantrene</td>
<td>4500</td>
<td>3.6</td>
<td>&gt;103</td>
<td>&gt;61</td>
</tr>
</tbody>
</table>

Figure 3.3: Experiment Data

In Figure 3.3, on the leftmost column, Kallmann listed the variety of materials he experimented on as of 1949. He hit the materials with varying levels of intensity of alpha ($\alpha$), beta ($\beta$), and gamma ($\gamma$) rays.

In the middle columns of Figure 3.3 were the two dimensions along which Kallmann evaluated candidate crystal materials. The first dimension is called “Physical Yield.” “Physical Yield” is a
measurement of how much of the incoming energy (alpha, beta, or gamma rays in Figure 3.3) is converted into visible light. Holding the incoming energy constant, the higher the physical yield, the brighter the visible light, and hence the bigger the pulse in Figure 3.2. For this reason, after he learned from experiments that Zinc Sulfide has around 27% of physical light yield and “approaches the maximum possible,” Kallmann wrote that “[t]his current can be easily measured particularly when the tube is cooled to reduce the dark current.” (Kallmann 1949, p. 624) In the context, “the dark current” referred to the background electrical currents already in the tube-shaped photomultiplier. When the currents were reduced in low temperature, 27% physical yield could provide a relative strong and easily recognizable peak in detection instruments.

The second dimension is called ”Technical Yield.” ”Technical Yield” is a measurement of how much of the converted light managed to escape the crystal. Based on the “physical yield,” Kallmann could already measure how efficient the crystal would be in converting radiation into light. But mere conversion was not enough. The converted, visible light also had to reach the photomultiplier. In the case of the gamma ray, Kallmann knew at the time that, other things equal, the opaquer and thicker the crystal was, the more likely it would be able to stop, capture, and convert incoming gamma rays into visible light. At the same time, the more opaque and thicker the crystal was, the less likely the converted visible light would be able to escape the crystal and reach the photomultiplier. If too much of the converted visible light was absorbed by the crystal, then the escaped light might not be bright enough to create a “strong” pulse. Therefore, Kallmann had to find a balance: he wanted a crystal that would turn as much of the incoming gamma ray into visible light as possible, while letting as much of the visible light escape as possible. In the end, Kallmann settled with Naphthalene as the best balance for radiation detection and used it as the crystal in his newly invented organic radiation detector.

Another important thing to note is that Kallmann also cared about sensitivity. Sensitivity is a measurement of how “wasteful” a detection instrument is. A detection instrument is more sensitive
if it can detect a higher percentage of useful information from the generally available pool of useful information. It is less sensitive if it picks up little from the generally available pool of useful information. If Kallmann’s organic radiation detector wasted most of the incoming radiation (such as by not converting them into visible light, or by absorbing the converted visible light), then it is a fairly low sensitivity system. In Kallmann’s case, one obvious way to increase the signal-to-noise ratio of his organic radiation detector was to increase sensitivity: given the fixed amount of energy of each incoming gamma ray, a straightforward way to increase the brightness of the converted flash is to use more of the fixed amount. Because the superior goal was signal-to-noise ratio in Kallmann’s experiments, I am focusing on it in this chapter. I will cover the scientific and philosophical significance of sensitivity in the next chapter.

Now that I have briefly explained Kallmann’s experimental concerns, let me take a pause, step back, and explain at a grand scale what Kallmann was trying to accomplish. His primary goal was not to validate or reject any particular theory. His primary goal was not to study the phenomenon of radiation or the phenomenon of scintillation either. He was trying to collect and identify a profile of physical, causal interactions: he shone scintillating materials with known, given amounts of radiation; he then measured how much visible light came out of scintillating materials. He was, in other words, trying to collect profiles of different materials, examine how well they reacted to radiation, and pick out the best candidate to become his future scintillator.

The discovery of the organic radiation detector, especially the organic gamma ray detector, happened at a fortunate time: precisely when nuclear physics as a field was looking for new directions. Kallmann’s invention of a high signal-to-noise ratio gamma ray detector, by making precisely detecting and counting gamma rays possible, enabled researchers, including researchers who came from other professions, to study nuclear physics. And nuclear physics was looking for new directions precisely at that time. Here, let me focus on developments in the United States. After World War II, the American government, research institutions, and researchers faced a public relations challenge.
During the war, the U.S. maintained a big nuclear research program, including the Manhattan Project, leading up to the first ever nuclear bomb. But when the war was over, all involved parties lost the primary reason to justify the continued existence of large-scale nuclear research (Creager 2013, p. 1). They had to find exciting civilian uses of nuclear power to justify their funding and projects. Among those transitioning away from military research and back into civilian discoveries was Martin Deutsch. Martin Deutsch was a renowned nuclear physicist and was a part of the Manhattan Project. Upon the end of World War II, he took up a professorship at Massachusetts Institute of Technology and started new projects on nuclear physics. He heard of Kallmann’s research from the American occupation forces in Germany (Martin Deutsch 1974, p. 68), submitted an abstract to an American Physical Society conference to describe Kallmann’s work (Deutsch 1948a), and translated and introduced Kallmann’s work into English (Deutsch 1948b). With Deutsch’s reporting, Kallmann quickly received letters and invitations from the United States, England, France, Russia, among other countries for his discovery of the organic radiation detector (Hine 1977, p. 870). At the same time, Kallmann already disliked post-War Germany, thinking that it never eradicated Nazi influence and was looking for ways to relocate to the United States (Charpa and Deichmann 2007, p. 259). Thanks to the newfound fame, he secured an invitation to develop his detector at the Army Signal Corps in Fort Monmouth in the U.S., before he was appointed a professor at New York University (Spruch 2008).

Because Kallmann’s organic radiation detector was precise, high signal-to-noise ratio, and easy to operate, the availability of the detector encouraged explorations of new research projects outside the traditional borders of nuclear physics. One potential direction for valuable, civilian nuclear physics research was and still is to use radiation to diagnose or treat tumors. Working somewhat independently, researchers in the U.S. embarked on two research directions that iteratively reinforced each other and ultimately led to scientific and practical successes in tumor diagnosis and treatment as we know it today. First, they wanted to improve upon Kallmann’s design and find better instruments
to detect gamma rays. Second, they wanted to use the best available detection instruments to study the properties of gamma rays or any nuclear phenomena involving gamma rays.

Concerning the first, the search for better crystals never stopped. Researchers were (and still are) invested in finding and manufacturing crystals that generate the biggest flashes. Among researchers who made significant contributions to the discovery of new and better crystals is Persa Raymond Bell. Bell finished his doctoral research at University of Chicago working on nuclear research. He was mobilized to work on radar systems at Massachusetts Institute of Technology during World War II. Upon the end of World War II, he returned to Oak Ridge National Laboratory to study nuclear fusion as an energy source. In the process, he was looking for a better way to detect and count radiation. In 1948, Bell confirmed that Anthracene was a better organic solid than Naphthalene, as proposed by Kallmann, to convert radiation into visible light (Bell 1948). Bell set up two sets of apparatus, one with Anthracene and another with Naphthalene. He produced Anthracene and Naphthalene crystals of roughly the same size. In both sets, Bell connected the crystals to a set of amplifiers to amplify the flash, and then to a pulse selector to automatically record the flash. Then, he started to hit Anthracene and Naphthalene with gamma, beta, and alpha particles. It turned out that, when hit with gamma particles, Anthracene produced “pulses” three times bigger than Naphthalene. Anthracene was not too efficient in responding to beta particles. But it was very efficient in capturing alpha particles – pulses five times bigger than Naphthalene. Because Anthracene was able to generate bigger flashes, which in turn stand out better from the background, the new detector delivered a much higher signal-to-noise ratio.

Similar to Bell, Hofstadter is another former World War II nuclear researcher who contributed to the discovery of even better crystals. Hofstadter discovered the Sodium Iodide doped with Thallium detector, often abbreviated as the NaI(Tl) detector (Hofstadter and Stein 1975). Hofstadter got his PhD during the interwar years and had primarily worked in basic physics. He was recruited to work on nuclear research during World War II. But upon the end of the war, he returned to basic physics.
He happened to be working on KI(Tl), when he heard about Kallmann’s work. He immediately tested KI(Tl) and found it to be roughly on par with Naphthalene and Anthracene. Going through the literature, he saw some papers suggesting that NaI could be a very good scintillating material, although no experiment had been performed. So he ordered some NaI, produced NaI(Tl), and followed Bell’s procedure to experiment on the new crystal. It turned out that the new NaI(Tl) detector produced twice as large a pulse as the Anthracene detector.

While Hofstadter was primarily searching for a crystal with bigger flashes, Hofstadter also realized that the new crystal would work better for gamma rays. He noticed that, for reasons he could not fathom then, radiation at a higher energy level tended to produce smaller flashes than radiation at a lower energy level on the same crystal (Hofstadter 1949, p. 796). In other words, detectors working very well with alpha rays would produce much dimmer flashes with gamma rays. NaI(Tl) was able to produce bright flashes for high energy gamma rays, making it a rare, ideal crystal for gamma rays. NaI(Tl) was not completely ideal, because each flash took 230 nanoseconds. In other words, it is possible that, when the next gamma ray hits, NaI(Tl) would still be flashing from the previous hit (Hofstadter 1949; McIntyre and Hofstadter 1950). As NaI(Tl) became the standard crystal for gamma ray detection, researchers for decades had to work around its shortcoming: future generations of gamma ray detectors would not be able to capture gamma rays fast enough or precisely record when they were detected without either changing the crystal or modifying other components.

So far, I have covered one research direction on finding the better crystal. The pattern in this research direction is straightforward. Following Kallmann’s practice, researchers tried to identify organic materials that would produce bigger and brighter flashes out of incoming radiation. They did this because they were pursuing after a better signal-to-noise ratio. They were able to do this, echoing Chapter 2, partially because the organic radiation detector was modular. They could identify better crystals and replace existing crystals in the detector, without needing to completely redesign other parts, such as the photomultiplier and the camera. The modular nature of the detector enabled
researchers to quickly improve their system along one dimension (in this case, signal-to-noise ratio), without worrying too much that improvement in one dimension may induce regressions in others.

At the same time, I observe the constant shifting of priorities among norms and preferences. Sometimes researchers craved crystals to better detect gamma rays, instead of, say, alpha rays. Sometimes researchers wanted bigger flashes in exchange for slower temporal response. In this chapter, I am merely making an observation of the history. In two chapters, I will philosophically investigate how the priorities of those norms and preferences can shift due to practical matters, in ways not unlike an evolutionary process.

### 3.2.2 Coincidence Detection and Adoption of Gamma Ray Detection in Medicine

At the same time, researchers were also keen to apply their new detection technology to study unknown phenomena. One direction was to rely on gamma rays’ penetrative nature to study the inside of human body, especially the brain, without physically opening up the body. For example, if we can precisely place a radioactive source inside a brain tumor and detect radiation from the outside, we will be able to locate the brain tumor without opening up the skull. In grappling with this opportunity, researchers quickly ran into another signal-to-noise challenge (Prince and Links 2014, p. 319): scattering.

Figure 3.4 is illustrates the challenge. Let us assume that we place gamma sources inside the skull at points A, B, C, D, and they are detected by a ring of detectors on the outside. Gamma rays will be scattered by tissues, the skull, and other radiation on their way out of the skull. In the case of point C, soft tissues scattered gamma rays and they made a “turn” on their way out. In the case of point B and point D, gamma rays collided with each other and turned each other astray. Only in the case of point A did the gamma rays travel without scattering or inference. But in this case, one detector did not flash in response to incoming A radiation – as illustrated by a shade in the figure, because the incoming angle was too narrow. Hadd point A registered with two detectors on both ends, researchers then could correctly and easily infer where the gamma rays originally came
Figure 3.4: Scattered Gamma Rays

from: they would note two data points at the opposite ends of the ring of detectors and infer that the gamma source was along the straight line between the two data points. In all other cases, when researchers looked at data points at the opposite ends of the ring and infer B, C, D were somewhere along the straight line between the data points, they would mis-locate those points. At the same time, researchers did not have other good ways to model the complicated scattering relationships for B, C, or D to avoid mis-location.

Fortunately, researchers came up with a solution to separate data points from unscattered gamma rays (signal) from those from scattered gamma rays (noise) with a technology called Coincidence
Detection. This section briefly covers the original history of Coincidence Detection, which will play a bigger role in the next chapter.

The basics of Coincidence Detection were possible partially because of another former Manhattan Project member. I have already discussed Martin Deutsch earlier in this chapter. Martin Deutsch, after his time at the Manhattan Project, returned to the Massachusetts Institute of Technology, helped to strengthen the department by recruiting his friends, and educated the next generation of Nobel prize winners in physics. Deutsch was not alone in building up departments through his connections and influences during the Manhattan Project years. In 1946, Washington University in St. Louis recruited the renowned physicist and Nobel Laureate, Arthur Compton to become its ninth chancellor, after he completed his work with the Manhattan Project. Washington University in St. Louis wanted Compton to build up not only its physics department, but also the university’s general research capacities, including attracting talents and hopefully managing and operating the Los Alamos National Laboratory, as the national laboratory was transitioning out of the Manhattan Project. Compton quickly recruited his fellow nuclear physicists during the Manhattan Project years and solidified the U.S. government’s support for nuclear physics research at Washington University in St. Louis. Among the list of nuclear physics breakthroughs at Washington University is the discovery of the angle of gamma ray emission upon positron annihilation by DeBenedetti et al.’s research (DeBenedetti et al. 1950). In DeBenedetti et al.’s research, they used the Anthracene detector to study the phenomenon of positron annihilation. It was known that, upon annihilation, a positron emitted a pair of gamma rays (Yang 1950). But it was not clear how the directions of the two emissions would be relative to one another. DeBenedetti et al. set out to find out the angle. They manufactured a positron source and used a pair of Anthracene detectors. They moved the pair of detectors around the source and recorded how many gamma rays the detectors captured at each angle. Hence, they were able to tell at which angle the readouts from the pair of scintillators abruptly peaked.
Figure 3.5: DeBenedetti et al.’s Qualitative Determination of the Angle

On top of Figure 3.5, DeBenedetti et al. sketched the experimental setup. They had the positron source in the center (S). And one detector on each side (D1 and D2). They moved one detector around at various angles from the horizontal line (moving D1). In Figure 3.5, the curve depicted detected counts of radioactivity. Radioactivity peaked between -1 degree and 0 degree away from the horizontal line. In other words, DeBenedetti et al. found out that the emitted gamma rays traveled almost precisely in opposite directions.

Other researchers quickly realized that DeBenedetti et al.’s research might allow them to increase the signal-to-noise ratio for positron detection. Wrenn et al. were among the first to point out the instrumental value of combining scientists’ knowledge of positron annihilation and the new NaI(Tl) detector. Frank R. Wrenn, at the time, was at the frontier of the post-war generation to apply nuclear research into actual medical practice. After his M.D., Wrenn received an Atomic Energy Commission fellowship and worked at Duke University to apply nuclear research to the localization of tumors. Wrenn et al. thought that they could greatly improve signal-to-noise ratio, sensitivity, precision and other data generation features by detecting positron annihilation through Coincidence Detection. Coincidence Detection means that the computer registers a “hit” only when a pair of
gamma rays hit a pair of detectors at the same time \(^3\) in almost the opposite direction. Each hit means that some positron annihilated along the straight line between where those two gamma rays arrive, also called a \textit{line of response}.

Coincidence Detection improves signal-to-noise ratio because scattered gamma rays are unlikely to arrive simultaneously at 180 degrees (Wrenn, Good, and Handler 1951).

- In the event that a positron emission takes place and the resulting two gamma rays leave the skull unscattered, they will hit two detectors at roughly 180 degrees at roughly the same time. The system will register a hit.

- In the event that a positron emission takes place, but the resulting two gamma rays are scattered on their way out, they will most likely not hit two detectors at roughly 180 degrees or at the same time. The system will NOT register a hit.

- In very rare occasions, two scattered gamma rays from different sources may happen to arrive 180 degrees apart at the same time. The system will register a hit.

If the source of positron emission event was not precisely at the center of the ring-shaped detectors, the emitted gamma rays would arrive at the two detectors at slightly different times. But given that the radiation traveled at the speed of light, and that Na(Tl) was a slow responding crystal, minor timing differences as such were not really detectable in the 1950s. Researchers eventually came up with a way to detect such minor timing differences and were better able to locate where the position emission event took place, but it took them another twenty years. I will cover that part of history in Chapter 5. In short, the system would register a hit if unscattered gamma rays from the same positron source arrived in 180 degrees at the same time (signal) and if superfluous coincidence happened when two unrelated gamma rays arrived at the same time in opposite directions (noise).

\(^3\) Gamma rays do not actually arrive at the same time because, unless the positron source is precisely at the center, equal distance to two detectors on the side, the pair of gamma rays will travel different distances and hence take different time to reach the two detectors. However, due to the slow response of NaI(Tl) detectors, researchers really could not tell the temporal difference.
However, otherwise scattered gamma rays would not register a hit in the system; hence, this greatly reduce the amount of noise recorded by the system. Given the very low probability of superfluous coincidence, the system achieved a very high signal-to-noise ratio.

For a comparison and to better explain, I have made a demonstrative table: “Signal” refers to useful information picked up by the detection instrument. “Noise” refers to useless information picked up by the detection instrument. “Signal-to-noise ratio,” roughly speaking, is the percentage of “signal” over “noise” in all the information picked up by the detection instrument. In Table 3.1, with or without coincidence detection, the detected amount of signal is roughly the same. Without coincidence detection, the unscattered gamma rays are picked up and are considered signal. With coincidence detection, only the simultaneously arriving gamma rays are considered signal. Only in a few exceptional cases, such as case A in Figure 3.4 where the gamma ray arrives at too narrow an angle, will the coincidence detection system miss the signal. But the detected amount of noise is significantly reduced with coincidence detection. Without coincidence detection, all the scattered gamma rays are picked up as noise in the system. With coincidence detection, only the superfluously coincidental gamma rays, two gamma rays coming from different sources but incidentally arriving at the same time in 180 degrees, are picked up as noise. In other words, the amount of signal detected is roughly the same with or without coincidence detection, but the amount of noise picked up is significantly reduced with coincidence detection. In total, coincidence detection improves signal-to-noise ratio.

<table>
<thead>
<tr>
<th></th>
<th>Without Coincidence Detection</th>
<th>With Coincidence Detection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal</td>
<td>Unscattered Gamma Rays</td>
<td>Most Unscattered Gamma Rays</td>
</tr>
<tr>
<td>Noise</td>
<td>Scattered Gamma Rays</td>
<td>Superfluously Coincidental Arrivals</td>
</tr>
</tbody>
</table>

Table 3.1: Signal versus Noise with and without Coincidence Detection
By around the mid-1950s, multiple teams had developed both good detectors to capture gamma rays and the mechanism to utilize coincidence detection for high signal-to-noise detection of positron emission. While Wrenn et al. worked out the basic mechanism of coincidence detection, Brownell’s team at Massachusetts General Hospital discovered coincidence detection independently at the same time (Sweet 1951). Now, researchers had the basic ingredients of a technology that eventually became Positron Emission Tomography. I will further explain the process of how those basic ingredients eventually evolved into Positron Emission Tomography in future chapters. In this chapter and in the following sections, my next goal is to reflect for a moment on the scientific and philosophical significance of increasing signal-to-noise ratio.

3.3 Philosophy, Experiments, and Noise

With the early history of PET covered, I dedicate this section to discussing the philosophical significance of signal-to-noise ratio in scientific instrumentation. I begin with a review of the literature to demonstrate the philosophical distinctiveness of signal-to-noise ratio for at least three reasons:

- It is a measurement of detection instruments and data generation, independent of particular theories to be tested. It does not treat detection instruments as an invisible process taken for granted, but instead at the center of data generation.

- It is diagnostic and offers a way to mechanistically improve the causal design of instruments for better signal-to-noise ratio.

- It offers a partial solution to Woodward’s problem of inference.

I then offer a scientific and philosophical account of signal-to-noise ratio.
3.3.1 The Links Between Data and Phenomena

Despite detection instruments’ central role in producing data from natural phenomena, many philosophers skip this critical component that links data to natural phenomena of interest to them. In this subsection, I discuss two common ways to neglect detection instruments in discussing scientific practices and data generation. In the first way, philosophers evaluate the epistemic value of data without assuming the specifics of detection instruments. They act almost as if data are just mysteriously presented to researchers, instead of being produced by specific and epistemically and practically constrained detection instruments. For this reason, those proposed epistemic virtues are often evaluative but not diagnostic: diagnostic virtues, like the ones I discuss in this chapter, often have concrete assessments of detection instruments, especially their causal mechanisms and offer clear paths to improving the instruments. In the second way, philosophers conceptualize signal and noise as features of data, without connecting them to features of detection instruments.

First of all, to understand and evaluate detection, contemporary philosophers of science commonly focus on at least two epistemic norms of detection and measurement, accuracy and reliability. Here, I quote Tal’s discussion of accuracy and Wimsatt’s discussion of reliability. Tal suggests that epistemic accuracy means the agreement between the attributed values and the actual value (Tal 2015, p. 1084). Here, “attributed values” are plural because contemporary measurement normally reports a range of possible values to indicate uncertainty. For example, if the actual temperature at my home is 73 degrees Fahrenheit, and if my thermostat reports that the temperature is between 72.8 degrees and 73.2 degrees, my thermostat is fairly accurate. The measurement is more accurate if the range of possible values is smaller and closer to the actual value. My thermostat would be more accurate if it reported that the temperature was between 72.99 degrees and 73.01 degrees. In Tal’s account, measurement is intrinsically about reducing uncertainty and error (p. 1084). Researchers interested in measurement and detection are primarily interested in reducing the range of uncertainty and errors as possible.
Wimsatt, in a recent systematic discussion of reliability and robustness, lists at least two reasons for the importance of reliability in measurement and detection. In his context, reliability refers to the invariant measurement results and/or causal mechanisms across multiple, independent determinations (Wimsatt 2007, p. 44). Reliability is critical in measurement and detection, first of all, because it reduces the risk of erroneous assumptions or causal mechanisms. Quoting Campbell, Wimsatt observes that the causal link between any measurement and the target phenomenon is often long and prone for actual or presumed irrelevant or non-ideal causal influences (pp. 53-54). Using multiple means of detection and measurement can potentially reveal the causally invariant link between measurement and target phenomenon and hopefully screen off the unwanted causal interference unique to each particular measurement. Secondly, reliability is often a good indicator that researchers are measuring something real. Wimsatt, again quoting Campbell and others, suggests that it would be a miracle (or an extremely unlikely situation) if we could detect and measure an object through multiple independent means even though the object did not exist (p. 60).

While accuracy, reliability, and other norms are extremely important in measurement and detection, they do not pay sufficient attention to detection instruments. Detection instruments are largely invisible in their discussion. In the case of Tal’s account of accuracy, accuracy is really about the results generated by detection instruments and other data-process methods. In the case of Wimsatt, it is more about using multiple detection instruments and other means of detection, than a practical guide to improve them. Those norms are not too useful for detection instruments partially because they are not diagnostic. They can evaluate, especially compare, between detection instruments. But they do not point out what is wrong or could be improved with those detection instruments. In the case of the thermostat, one might be tempted to ask what are the critical design features in the thermostat to reduce its reported range from $±2$ to $±0.01$? Without diagnostic features, the aforementioned epistemic virtues are very much similar to telling me that getting a B- was not good in a philosophy class: it is informative, but does not really tell me how I can do better. In this
chapter and throughout the dissertation, I discuss engineering and epistemic virtues that are not only evaluative, but also diagnostic.

To discuss the first such virtue – increasing signal-to-noise ratio, I review the definitions of signal and noise offered by Bogen and Woodward and McAllister. They offer competing definitions primarily because Woodward and Bogen and McAllister have a realism debate. Woodward and Bogen believe that natural phenomena exist as mind-independent components of the natural world, and researchers’ epistemic aim is to infer from observed data to those natural phenomena. McAllister does not think such mind-independent natural phenomena exist. He thinks that researchers stipulate patterns in data as natural phenomena and all natural phenomena identified by different researchers co-exist at the same time (McAllister 2010, p. 810).

Because of their differences on realism, they offer different ontological accounts for signal and noise. Woodward (and in agreement with Bogen (Bogen and Woodward 1988)) thinks that “signal” consist of data that are causally (and ideally) produced by the right kind of causal interaction with target phenomena, while “noise” consists of data that are causally produced by any additional, interfering causal interactions (Woodward 2010, p. 793). In general, McAllister think that “signal” or “phenomena” are patterns identified in empirical data by researchers. “Noise” then is the unexplained leftover from the empirical data (McAllister 1997, p. 219).

I will avoid the larger realism debate between Bogen and Woodward and McAllister and focus instead on their notions of signal and noise. I argue that while they offer valuable insights into the ontology of and roles of signal and noise in science and in philosophy, they neglect both the mathematical nature of signal and noise and the intimate relationship between signal and noise and detection technology. After all, noise (or signal) does not come from a vacuum. By tracking the causal mechanisms that create signal and noise, contemporary researchers are able to define them in mathematical and causal terms. Once we do have an account of the signal and the noise based on
detection technology, we will realize that such an account can break free from Bogen and Woodward
and McAllister’s debates on ontology and realism.

Bogen’s recent writings on this topic are primarily focused on challenging McAllister and teasing
out different notions of noise. According to Bogen, two senses of noise exist (Bogen 2010, p. 781).
One sense of noise consists of data points that interfere with data points of interest and useful to
us. Noise makes it harder for researchers to identify and utilize data points that are generated by
intended and ideal causal interactions with the phenomena of interest to them and is hence defined
relative to data points researchers want to collect. Bogen identifies another sense of noise that is not
relative to the phenomena or signal of interest to us. Data points can be noisy simply because they
are not well organized.

In this chapter, I focus only on the first sense of noise identified by Bogen as the “interfering” sense
of noise. A good example of the definition for the “interfering” sense of noise probably comes
from Data, Phenomena, Signal, and Noise, a paper by Woodward. In the paper, Woodward takes
signal and noise to be components of data. He thinks that, in general, signal and noise are generated
and primarily defined by desirable and undesirable causal interactions between data and natural

Investigators try to arrange data production processes so that the data are causally
influenced by the phenomena they are trying to learn about (and because of this convey
information about those phenomena), but commonly data will also reflect the causal
influence of many other more local and idiosyncratic factors having to do with properties
of instruments, measurement procedures, and the environment. The influence of these
additional factors on the data is often conceptualized as “error” or “noise.”

Woodward makes the following claims in the quote and passages around the quote:
• Ontology: Natural phenomena exist, independent of human interest, observation, and interaction

• Epistemic Goal: Researchers want to gather data about the phenomena of interest to them through ideal causal processes

• Nature and Source of Noise: Noise is the product of additional causal influence outside the right kind of causal interaction in the process of data gathering

While McAllister also talks about noise as components of data, McAllister systematically disagrees with Woodward (and Bogen) (McAllister 1997, 2010). McAllister’s challenge to Woodward seems to be a challenge of realism. McAllister does not believe that phenomena truly exist as mind-independent elements of the natural world. Instead, researchers have infinitely many ways to carve up data and divide them into patterns and noise based on their interests (McAllister 1997, p. 219). Signal and noise are merely researchers’ stipulation. In his earlier, 1997 paper, he offers the following account of phenomena and noise (pp. 223-224):

Investigators analyzing data sets encounter infinitely many patterns exhibited with various noise levels. Each investigator designates some of these patterns, picked out on any criterion that he or she may choose, as the patterns that correspond to phenomena. Far from denoting a small number of fundamental constituents of the world, the term ‘phenomenon’ is on my account a label that investigators apply to whichever patterns in data sets they wish so to designate. Thus, on my account, which patterns count as those corresponding to phenomena is entirely a matter of stipulation by investigators. Typically, the patterns that investigators regard as corresponding to phenomena are those that they intend to study or hope to explain.

It is worthwhile to make a parallel comparison between Bogen and Woodward and McAllister:
As demonstrated by Table 3.2, primarily focusing on the interfering sense of noise, Bogen and Woodward and McAllister have completely different theories about the ontology of noise, the epistemic goal of data processing, and the specific methods involved in data processing. At the same time, they share the assumption that their definitions of signal and noise depend on their positions on realism and ontology. In the next section, when I offer my account for the signal, the noise, and the signal-to-noise ratio, I offer an account largely independent of realism debates. Perhaps it is the obsession with realism that hampers the independence of philosophy of experimentation.

### 3.3.1.1 The Problem of Inference

During Woodward’s discussion of signal and noise, he also points out that a critical epistemic challenge in science is inference from observed data to target phenomenon of research interest:

> These features create a generic epistemic problem: investigators must somehow use data that is noisy and, where the details of the process producing this noise are not fully understood, learn about the phenomena of interest (Woodward 2010, p. 794).

Researchers almost never directly observe most of the phenomena they intend to study: no one literally “observes” the melting point of lead, for example. Researchers rely on many and often complicated causal relationships to measure and transform information about the target phenomenon into observable data. And they infer back to the phenomenon based on their observable data.
Woodward is by no means the only person pondering this challenge but is perhaps more systematic than others. Chang (Chang 2007, p. 43) and Roskies (Roskies 2008) note similar challenges in inferring from data to target phenomenon. But Woodward goes further to offer a mathematical and systematic formulation for this epistemic challenge:

\[ d_i = f(P, u_i) \]

In the formula, \( d \) refers to the observed individual data points. \( P \) refers to the phenomena of interest and \( u \) consists of additional causal factors. \( f \) is the function that connects natural phenomena and additional causal factors to data points. Woodward assumes that one epistemic goal of using data is to infer back to the phenomena (Woodward 2010, p. 795):

[The researcher’s] goal is to use the \( d_i \) to make an inference about something else:

\[ P \]

The existence of noise makes the goal difficult to achieve: researchers have only very incomplete knowledge of the function \( f \), \( P \), and \( u \). With only \( d \) in researchers’ hands and \( f \) largely unknown, researchers often do not have enough information to infer back to \( P \).

In this passage, Woodward implicitly opens the door to a solution to his problem. Although \( f \) is an all-encompassing function that includes all the connections between data and the target phenomena (and the noise), in many controlled environments, especially in experiments, the detection instrument composes a significant if not major portion of \( f \). Through profiling the detection instrument, researchers can have a pretty good idea of \( f \). One feature that researchers regularly profile is signal-to-noise ratio.

### 3.3.2 A Scientific and Philosophical Account of Signal-to-Noise Ratio

Following Bogen, Woodward, and McAllister’s discussions, I intend to offer an alternative account of signal, noise, and signal-to-noise ratio in an instrument-centric context. In this instrument-centric context, I foreground the significant role of detection instruments in generating scientifically valuable
data, and background the traditional philosophical discussions of theory-choice or theory-ladenness in observation, etc. I argue that signal-to-noise ratio is a detection instrument centric measurement, has diagnostic features of detection instruments, and it partially addresses Woodward’s problem of inference.

3.3.2.1 Defining Signal-to-Noise ratio

Definitions of signal, noise, and signal-to-noise ratio are recurring and early themes of common textbooks in signal detection, imaging, and signal processing. Authors of such textbooks have in general agreed on basic features of noise and signal-to-noise ratio. Unlike McAllister, they do not define noise merely as a part of data, especially not as the unexplained part of data not included in the “pattern.” Similar to Bogen and Woodward, they define noise as caused by additional and unwanted causal relationships. But, by signal-to-noise ratio, they refer more to features of detection instruments than the generated data. They define signal and noise not as components of data, but as features of detection instruments. Researchers have generally agreed that signal and noise are defined based on their inferential value. The following is a quote from one textbook on medical imaging:

An unwanted characteristic of medical imaging systems is noise. Noise is a generic term that refers to any type of random fluctuation in an image, and it can have a dramatic impact on image quality; image quality decreases as noise increases (Prince and Links 2014, p. 69).

Let us review another quote from a general signal processing textbook:

In communication systems, message signals are corrupted by unwanted signals (noise) during transmission over a channel. The quality of the received signal is judged by the relative sizes of the desired signal and the unwanted noise. In this case, the ratio of the message signal and noise signal powers (signal to noise power ratio) is a good indication of the received signal quality (Lathi and Green 2014, p. 25).
The sensitivity and accuracy of any detection system is limited by random fluctuations that always accompany the measurement. It also sets a limit to the minimum detectable signal. These random fluctuations or disturbing signals, called noise, [...] (Witteman 2006, p. 1).

I must note that “random” in technical terms does not mean metaphysical randomness. “Random” here refers to epistemically unknown and unexplained patterns. In short, those quotes converge on a few features of noise:

- Noise is an undesirable and integral part of the data-generating feature of detection instruments (“characteristic of [...] imaging systems,” “[i]n communication systems,” etc.)
- Signal-to-noise ratio is a good indicator of data and detection instruments’ quality
- Increasing signal-to-noise ratio makes signal more discernible and recoverable.

While the description of noise may sound like Woodward’s definition, textbook authors are talking about intrinsic properties of detection instruments, instead of features of data. Let me make a comparison. Earlier in this paper, I mentioned Woodward’s mathematical formula for data:

\[ d_i = f(P, u_i) \]

In his formulation, data are a function of the genuine phenomena \( P \) and other interfering causal factors \( u \). Those interfering causal factors \( u \) are the source of noise in data. The above textbook authors are talking about properties of detection instruments. The following is a formula from one signal-processing textbook (Prince and Links 2014, p. 76):

\[ G = f + N \]

In this book, we often assume that the output of a medical imaging system is a random variable \( G \) (or a collection of random variables), composed of two components, \( f \)
and $N$. Component $f$, which is usually referred to as signal, is the (deterministic or nonrandom) “true” value of $G$, whereas $N$ is a random fluctuation or error component due to noise. The identification of an abnormal condition within the human body most often depends on how “close” an observed value $g$ of $G$, characteristic to that condition, is to its true value $f$. (Italics original)

The wording requires some deciphering. Here, $G$ is the output of some detection instruments. $G$ is basically the data, or $d_i$ for Woodward. It is the result of causally interacting with signal ($f$) and noise ($N$). So far, the description sounds similar to Woodward.

But the researchers are talking about an entirely different challenge: Woodward is interested in inferring from known $d_i$, or the data researchers have in mind, to $P$, or the natural phenomena that researchers are actually interested in. Researchers in the quoted passages are interested in figuring out the actual profile of their detection instruments: given ideal and known input such as $f$, what will come out of the detection instruments such as $G$? Once researchers understand the profile of their detection instruments, they can approximate the $f$ in Woodward’s sense, and along with $d_i$, infer $P$.

In other words, while we cannot directly observe the melting point of lead, as Woodward suggests, we can measure the signal-to-noise ratio of a detection instrument by feeding it known inputs. In the example of Kallmann, he fed many candidate scintillator materials with known amounts of alpha, beta, and gamma radiation. He then recorded how much of the radiation was converted into light, and how much of the light eventually escaped candidate materials. The process is similar to what Allan Franklin refers to as calibration: the process of feeding instruments with known inputs, examining whether instruments reproduce known phenomena in the right way, perhaps tinkering with them so they reproduce known phenomena, and then adjusting one’s beliefs in the epistemic value of data produced by those instruments (A. Franklin 1998, p. 237). The difference between Franklin’s and my account is that Franklin primarily focuses on the evidential value. Calibration,
according to Franklin, gives researchers an idea how good the results are as evidence for the inputs and target phenomena. My account here focuses on profiling the equipment: researchers intend to have a systematic understanding of features of the detection instruments (especially how they interact with target phenomena) and possibly identify ways to improve those features. The focus is on the instruments.

So far, I have provided some conceptual understanding of noise and signal-to-noise ratio. Now, I want to list a few ways that researchers present signal-to-noise ratio in mathematical forms. Because signal and noise can take many forms, researchers have proposed many different forms of signal-to-noise ratio.

Perhaps the most intuitive mathematical characterization of signal-to-noise ratio (SNR) is the amplitude SNR (Figure 3.6).

![Figure 3.6: Amplitude SNR](image.png)

Amplitude SNR assumes that, although signal and noise may overlap, they are more or less separate, and signal differs from noise by numeric differences in one or more dimensions of measurement. This is the type of SNR that Kallmann was dealing with, referring back to Figure 3.2. In Figure 3.6, the signal differs from the noise by having a much higher audio amplitude. But the difference does not have to be audio amplitude strictly. Amplitude SNR is an intentionally vague terms that
allows researchers to define amplitude in specific contexts (Prince and Links 2014, p. 77). Insofar as researchers can differentiate signal from noise numerically along some dimension, researchers can measure the amplitude SNR by

\[
SNR_{amplitude} = \frac{Amplitude(\text{Signal})}{Amplitude(\text{Noise})}
\]

However, in many other cases, researchers cannot differentiate signal from noise merely by numeric values along some dimensions. They might be able to differentiate signal from noise based on the “power” or “strength” and can characterize SNR through power SNR (p. 77). Especially in cases where the signal and the noise are additive, that they do not merely overlap, but may combine or influence each other in a significant way, using mere numeric differentiation would not work. Therefore, researchers have to resort to a different measure, the measurement of “power.” Power is, in short, a measurement of variation. It measures how the actual data deviate from the ideal signal to determine the extent to which the signal may be recoverable (p. 78). Let us review an example.

![Figure 3.7: Signal Only](image)

In the example, in Figure 3.7, researchers have the pure, ideal signal. In Figure 3.8, this signal is influenced by noise. Because researchers are measuring the property of instruments by feeding them with known phenomena, researchers do have a good understanding of the pure, ideal signal. In Figure 3.8, researchers can measure SNR by measuring the extent to which the variance muddies the trend. In Figure 3.8, because the noise is limited, it is still roughly possible to recover the original
However, in some cases, the variance may be so severe that researchers cannot recover the ideal signal anymore. For example, in Figure 3.9, because the noise is so bad and the variance from signal is too big, it would be impossible to recover signal merely from the data points. Hence power SNR here is measured by the integration of signal divided by the variance. Or in more general terms

$$SNR_{Power} = \frac{Power(Signal)}{Power(Noise)}$$

Of course, there are other forms of SNR measurement. Sometimes researchers are more interested in the difference between signal and noise (for example, in cases where researchers have universal background white noise, etc.) and they have come up with differential SNR (Prince and Links 2014,
But the general spirit of SNR should be that researchers define SNR to measure the extent an instrument can generate signal over noise and the extent to which a detection instrument allows for recovery of the original signal (and hence inference back to the original phenomena of interest to researchers). It is perhaps interesting to point out that decibel (dB), a common unit we use when talking about audio instruments, is in fact a standard SNR measurement. While the folk commonly thinks of decibel of a measurement of “loudness,” it in fact measures the extent signal (including the loudness of sound) stands out of noise. To be specific, decibels can be used to measure different kinds of SNR as mentioned in this section. If we are measuring amplitude SNR, then formula would be (Prince and Links 2014, p. 80):

$$\text{SNR (in dB)} = 20 \times \log_{10} \left( \frac{\text{Amplitude(Signal)}}{\text{Amplitude(Noise)}} \right)$$

If we are measuring power SNR, then the formula would be:

$$\text{SNR (in dB)} = 10 \times \log_{10} \left( \frac{\text{Power(Signal)}}{\text{Power(Noise)}} \right)$$

The presence of many forms of SNR measurement is philosophically interesting. Because signal and noise are defined according to researchers’ goals and interests, and because researchers intend SNR to provide diagnostic information, they come up with multiple forms of SNR to address different kinds of signal and noise relationships and different causal mechanisms.

### 3.3.2.2 Addressing the Problem of Inference

I have noted that signal-to-noise ratio is one measurement of the “function” that transforms ideal signal and natural phenomena into data in researchers’ hands. Knowing more about such a transformation function constrains and informs the reverse inference from data in researchers’ hands back to natural phenomena. Signal-to-noise ratio is intended as a partial fix to Woodward’s problem of inference.

I want to illustrate this point with a concrete example: the determination of the Point Spread Function of a detection instrument (p. 56).
In the Figure 3.10, researchers give a known “point source” (think of it as a controlled minuscule source of input, such as a very small spot of visible light) as an input into the imaging system and observe the resulting image. They repeat the process multiple times to build a model of Point Spread Function (PSF), namely, a function to connect input to output, or a known feature of the imaging system.

\[ \text{output} = f(\text{input}) \]

Now that \( f \) is known, researchers can figure out the output if they know the input and they can reverse infer to the original input if they know the output.

3.3.2.3 Strategies to Improve SNR

Now that I have discussed the definition of SNR, I move on to discuss a few strategies to improve SNR. Signal-to-noise ratio is important not only because it is a useful measurement of detection instruments’ performance. It is indispensable and distinct from other epistemic virtues such as reliability because it is also diagnostic. By diagnostic, I mean that the measurement can provide actionable information to evaluate and improve on the causal mechanism that it measures on.
In this chapter, I will not discuss data processing strategies to separate signal from noise. Bogen, Woodward, and McAllister have already discussed extensively methods and strategies to recognize patterns and signal from noise in data. I want to focus on improving SNR as an intrinsic feature of detection instruments.

First of all, the measurement of SNR gives a clear guideline for improving SNR. Let us review the amplitude and power formulas of SNR:

\[
SNR_{\text{Amplitude}} = \frac{\text{Amplitude}(\text{Signal})}{\text{Amplitude}(\text{Noise})}
\]

\[
SNR_{\text{Power}} = \frac{\text{Power}(\text{Signal})}{\text{Power}(\text{Noise})}
\]

Based on the formulas for SNR, the general direction for improving SNR is straightforward. Researchers could increase the amplitude of signal, decrease the amplitude of noise, reduce the variance of noise, etc.

However, this general direction does not inform researchers of actual design choices in constructing their detection apparatus. The historical examples I discussed have offered three general directions, the comparative method, the amplification method, and the theory-guided method, in the increasing order of researchers’ knowledge of and control over the causal, detection process.

### 3.3.2.3.1 The Comparative Method

The first method, assumed by Kallmann, is an extensive comparison across various designs and design choices. The comparative method is essentially an extended trial-and-error for improved SNR profile. It does not require much causal knowledge of how some signal is detected. Kallmann did not fully understand the causal interaction between radiation and scintillators. He did not and could not make many causal predictions of how scintillators reacted to radiation. He could only rely on extended permutation: hitting a wide variety of scintillators with alpha, beta, and gamma radiation, collecting the resulting SNR profile, and picking the best one. The direction was continued for a long time after Kallmann, by Bell, Hofstadter, and many others. Hofstadter did not have a
good causal understanding of scintillators’ SNR profile either. He tested radiation against KI(Tl) and NaI(Tl) and found a much better SNR profile in NaI(Tl).

The comparative method may look crude, but is in fact heavily constrained. It could look as if that researchers were almost aimlessly throwing darts and took whatever stuck. But the experience of Kallmann, Hofstadter, etc. showed that they were searching in a very constrained space, extending out from what is known and familiar. Kallmann, for practical reasons, had already determined that he needed an organic material, and he also knew materials similar to the content in his technician’s clothing would work. So he focused primarily on materials structurally similar to Coronene, the glowing content of the clothing. Similarly, Hofstadter was performing research on KI(Tl) at the time and he experimented on materials similar to his research target. In other words, given that researchers use the comparative method when they have limited ideas of what and why things work, they often rely on identifying materials that are similar to what has already been observed to work, and then systematically comparing them.

3.3.2.3.2 The Amplification Method

Secondly, without much causal understanding, researchers can sometimes still causally differentiate features of signal and noise, such as constructing a causal mechanism to magnify the difference. In this case, researchers still do not need to understand why or how the detection apparatus works in great detail. They only need to understand that signal and noise differ along some critical features. For example, let us say the needle in a haystack is 2 inches long while the average hay is 1 inch. Our naked eye can barely tell the difference of 1 inch at a distance. But what if we somehow amplify the size of both by ten times? Now the needle is 20 inches long, while the average hay is 10 inches. Telling the difference of 10 inches will not be too challenging. This is how the photomultiplier worked in Kallmann’s setup: by amplifying both the signal and the noise, minuscule differences become apparent and easier to process.

98
3.3.2.3 Theory-Guided Methods

Lastly, when researchers do have a good causal understanding of why and how signal and noise are produced in detection instruments, they can tailor their causal mechanisms to take advantage of the knowledge. This direction is best illustrated by the historical section on Coincidence Detection. Coincidence Detection is built on a fairly simple theoretical principle: unscattered pairs of gamma rays will arrive at about the same time 180 degrees from each other and register; scattered pairs of gamma rays will be unlikely to arrive at the same time and at 180 degrees from each other; incidental simultaneous arrival is very unlikely. Hence, Wrenn et al. constructed a detection apparatus that only registers a signal if there is simultaneous arrival of two gamma rays in 180 degrees.

3.4 Conclusion

In this chapter, on the historical side, I reviewed the early years of the invention of the gamma ray detector. The gamma ray detector is the product of many researchers’ ingenuity and historical coincidence. In a way, a major breakthrough came about when Kallmann noticed the glowing seams of his technician’s shirt. Following this discovery, Kallmann, Bell, Hofstadter, and many other researchers relied on the known phenomenon that Coronene could glow when energized by radiation and explored structurally similar organic materials to find the most suitable material for radiation detection.

Among their criteria for good radiation detection material was signal-to-noise ratio. In the history covered in this chapter, signal-to-noise ratio had a very specific meaning: researchers wanted to find an organic material that flashed the brightest against the background when hit by incoming radiation. Eventually, later generations of researchers integrated the gamma ray detector into the technology known as Positron Emission Tomography.

On the philosophical side, I offered a philosophical account of signal-to-noise ratio that is largely free of the realism and ontological debate by Bogen, Woodward, and McAllister. The realism debate
by Bogen, Woodward, and McAllister is focused on whether researchers’ purported phenomena are mind-independent or mere stipulations based on pattern detection in data. However, researchers covered in this dissertation barely wonder if gamma rays are real. Instead they focus their efforts on evaluating and improving their detection instruments – which are simply assumed to be real. Signal-to-noise ratio, under my account, is a measurable feature of detection instruments. It measures, on a high level, how well detection instruments solve the epistemic challenge of generating data that allow for easy separation and recovery of signal out of noise. It is measured by feeding known signals into detection instruments and mathematically profiling how known signals are transformed into data generated by detection instruments.

Signal-to-noise ratio is important for epistemology and philosophy of science for at least three reasons. First, it is an instrument-centric measurement. It does not assume that detection instruments are designed for testing specific theories. Nor does it assume that detection instruments are invisible in philosophical debates: as many quotations in the Introduction and in this chapter talk about the phenomena in nature and data we have, but simply skip the instruments and causal mechanisms in between.

Second, signal-to-noise ratio is concrete and diagnostic. It not only measures the performance of detection instruments, but also offers ways for improvement. In this case, I have considered the comparative, the amplification, and the theory-guided methods to improve signal-to-noise ratio.

Third, signal-to-noise ratio is a partial fix to Woodward’s problem of inference in science. Woodward points out that researchers often have data in their hands, but have extremely incomplete information on how their target natural phenomena are transformed into the data in their hands. But measuring how detection instruments transform natural phenomena (input) into data (output) offers exactly the much-needed critical information of the transformation process. The PSF is such an example: through calibration, researchers can collect how the detection instrument transforms a point source
into data and identify the transformation function. Knowing the transformation process constrains and informs the reverse inference from generated data to target natural phenomena.

The conceptual and normative discussion of signal-to-noise ratio is merely the first chapter on a general project to explore aspects of scientific detection and scientific instrumentation. Researchers construct detection instruments to separate, extract, and recover information. Researchers almost always need to separate useful, relevant information from irrelevant and interfering information and to distinguish between two pieces of relevant, qualitatively similar information. But extracting signal from noise is merely one goal for researchers. In the ensuing chapters, I will cover other goals and engineering virtues reflected in the history of PET and make a case for an instrument-centric view of philosophy of science.
Chapter 4

1950-1970s, The Invention of Positron Camera and the Epistemic Value of Sensitivity

4.1 Introduction

In the previous chapter, I discussed the role of increasing signal-to-noise ratio in information extraction and data generation. This chapter is a continuation of the previous chapter: it picks up the historical narrative left hanging at the end of the previous chapter and covers another norm in engineering detection instruments: the norm of increasing sensitivity.

I hope, by covering two decades of history and two unique engineering norms, I am providing a concrete philosophical picture of the independence of technology-driven scientific change and engineering norms. Technology-driven scientific change does not mean theories play no role in research or scientific development in general. Technology-driven scientific change means that theory-testing is not the main or solo driving force in many episodes of scientific change. In those
episodes, the history of Positron Emission Tomography being one of them, it is the availability of technologies, materials, pragmatic considerations, and engineering norms that drive scientific change. In Kallmann’s experience, the driving force of developing a scintillator was not to test his specific theories of the relationship between the chemical structure of crystals and radiation. The driving force included the pragmatic need to measure radiation, the goal of increasing signal-to-noise ratio, the coincidental discovery of one suitable material, and the extended trial-and-error process to find similar materials. In this chapter, I will focus on the history in which researchers, now with better instruments to measure radiation, wanted to find a marketable civilian applications for radiation measurement. They zigzagged across many fields, including but not limited to cancer treatment, cancer detection, live biological processing detection, etc. When they were looking for new directions and when they did not have many specific theories to test, they were driven by pragmatic concerns (diagnosing and treating cancer, for example), technologies (the radiation detection instruments at hand), and the governing engineering norms such as those of increasing signal-to-noise and increasing sensitivity.

Those engineering norms play an enormous role in technology-driven scientific change for a number of reasons. First, they inform researchers of how they can develop good detection instruments, in general. While those norms often have pragmatic trade-offs against one another – for example, increasing temporal resolution often means decreasing spatial resolution, each of the values is often a general desideratum in engineering and in designing detection instruments. When researchers struggle to find pragmatic uses for their instruments, they can always focus on building instruments that are better in general. Even when those norms are in conflict, they can focus on one or more compatible norms to improve on their instruments. Second, those norms are based on underlying measurements of instruments and can provide guidance on how researchers can improve their instruments. Both signal-to-noise ratio and sensitivity are causally informed and mathematical formula-based measurement of the performance of instruments. They clearly indicate which variables should be changed to causally improve the measurement.
But most importantly, those engineering norms are living examples of an instrument-centric perspective in understanding information and evidence. The instrument-centric perspective can be summarized in two sentences:

- It is not the case that instruments are good because the data they generate are good.
- Instead, the data they generate are good because the instruments, according to the engineering norms, are good causal mechanisms built to satisfy those engineering and detection norms.

As Bogen and Woodward argue, researchers most often do not observe target natural phenomena directly (Bogen and Woodward 1988). No one literally observes the melting point of lead. Instead, researchers collect data from detection instruments and infer back to target phenomena (for Bogen’s discussion of this and how this can cause problems in interpreting fMRI data, see: Bogen 2002). Given that researchers need to infer from data to phenomena, one wonders how researchers, in practice, assess the accuracy, precision, and other epistemic values of data, when phenomena are still unknown and waiting to be inferred and reasoned about. In practice, and as the history of Positron Emission Tomography tell us, researchers often anchor the quality of data and evidence based on the quality of their detection instruments. Data are good, trustworthy, accurate, and worthy of admission into theoretical reasoning, because they are the product of good detection instruments. But what make detection instruments good? “Good instruments” are those that fulfill measurement and engineering norms such as increasing signal-to-noise ratio and increasing sensitivity.

Thanks to those norms and to an instrument-centric perspective, researchers can be sure that data and good and are admissible into an evidential relationship because the instruments, the causal mechanisms that produce the data, are good. Data are not always in doubt, because the causal mechanisms that produce them are not in doubt. For sure, instruments themselves have to be evaluated, but often through a highly specific process: by feeding them with known information and measuring different aspects of how the information is processed by detection instruments – through
the measurement of signal-to-noise ratio, sensitivity, etc. Once the calibration and validation are
done, the quality of detection instruments becomes a good and widely relied on, although fallible
indicator of the quality of data generated by those instruments.

To further demonstrate the case of an instrument-centric perspective, this chapter consists of two
sections. The first section is the historical section, focusing on the development of Positron Camera
around the 1950s at Massachusetts General Hospital (MGH). In this part of history, I focus on the
role of sensitivity, as an engineering norm, in driving the development of Positron Camera. I then
devote one section to a philosophical definition and an analysis of sensitivity and its associated
norm, increasing sensitivity.

4.2 The History of Positron Camera

In the previous chapter, I stopped my history section when researchers started to realize that the
phenomenon of positron emission could become a useful means of medical detection in the 1950s.
In this chapter, I pick up the history from the 50s and focus on how a team of researchers turned the
possibility into reality. Their goal was not merely to put together a machine: as for most engineering
and application-oriented designs, the design of a device and the identification of its application go
hand-in-hand. The team had to create a device that actually works, known as Positron Camera – a
predecessor to Positron Emission Tomography, and find a medically useful venue to use the device.
The success recorded in this chapter is limited, as researchers never saw widespread interests in
their Positron Camera during the phase.

But such a history of largely unsuccessful attempts to develop a detection instrument with widespread
academic adoption offers valuable insight for philosophers. In this part of history, researchers were
not driven by particular research goals or specific applications. They were searching, in a somewhat
vague sense, for a useful medical application of their new technology and new device. The search
space was not well defined, although certainly limited by researchers’ own interests and experience
(in this case, brain tumor surgery). What then drove researchers’ search when their goals were vague? One driving force is a shared understanding of what makes a detection device good in general. Those general, normative values limited the search space and helped researchers clarify exactly what kind of technical improvements and directions they should pursue, even when a specific application was vague. In this part of history, I especially emphasize that it is the norm of increasing sensitivity – the general value to make more efficient uses of useful information – that drove researchers at Massachusetts General Hospital to develop their Positron Camera.

The history section of this Chapter is primarily focused on the research at Massachusetts General Hospital (MGH) under the leadership of Gordon Brownell. Brownell was trained as a physicist and conducted military-related research on deep-sea mine detection during World War II. After the War, he worked with William H. Sweet, a neurosurgeon at MGH on medical detection. Sweet often operated on brain tumors and influenced the direction of research to focus on tumor localization. Thanks to their research and invention, after much zigzagging, researchers created a new instrument known as Positron Camera and a new way to use radiation for medical imaging – for tumor localization.

The history section is divided into four subsections. The first subsection concerns efforts at University of California, Berkeley, in the 1930s, before the World War II, to use radiation to treat tumors. The second subsection concerns with early efforts at MGH, under the leadership of Brownell, to adopt discoveries at Berkeley to detect brain tumors. The third subsection concerns with finding new directions and theoretical principles of a new detection system prototype. Finally, the last subsection concerns with the eventual invention of Positron Camera.

4.2.1 The Discovery of Phosphorus-32’s Selective Concentration in Tumors

While the focal history of this chapter starts in the 1950s, the key insight to connect radiation and tumor detection came from the 1930s, with the discovery of X-ray. By 1950s, researchers were not foreign to using radiation for medical detection. X-ray technology was already fairly popular in the 1930s. But X-ray does not really distinguish tumors from normal soft tissue. Separating
tumors from normal tissue, under X-ray, relies on an indirect chain of inferences: tumors often cause unnatural deformation of internal structures (such as bone fractures, organ swelling, etc.). Based on those abnormal structural changes, visible under X-ray, radiologists could infer that the nearby tissue may be or contain tumors. But such inferences are not always reliable: tumors are far from the only cause of structural anomalies. Researchers at the time were hence looking for new methods to detect tumors.

New ideas for tumor detection came from a failed clinical trial in the 1930s. John H. Lawrence, a pioneer of nuclear medicine working at University of California, Berkeley, led a team of researchers to explore whether radiation could be used to treat cancer. John Lawrence was trained as an M.D., but long had research interests in physics and applying nuclear physics in medicine. His interest was perhaps not incidental because his brother was Ernest O. Lawrence, who invented the cyclotron and got a Nobel prize for it, was a major figure at the Manhattan Project, and co-founded the Lawrence Berkeley National Laboratory and the Lawrence Livermore National Laboratory (both of which bear his name). It was also convenient that Earnest O. Lawrence was working at Berkeley too. The family tie allowed John first-hand and easy access to the most pioneering nuclear research and facilities at the time. John’s access included Phosphorus-32, a radioactive isotope that could be made by Ernest’s cyclotron. Lawrence and his team in 1939 noted that it was known at the time, when normal mice were fed with Phosphorus-32, a substantial amount of Phosphorus-32 would deposit in the bone. It was also known that X-ray, another form of radiation, could treat leukemia. They formed the hypothesis that substantial deposits of radiation in the bone could be used to treat leukemia. But the Lawrence lab found no evidence that oral feeding Phosphorus-32 was clinically effective in treating leukemia. There was no evidence to support a difference in the course of disease among lymphomatous mice with or without Phosphorus-32 treatment (Lawrence and Scott 1939; Lawrence et al. 1940).
But, in analyzing the tissue from these experiments, they discovered that Phosphorus-32 selectively concentrates at tumor sites.

Figure 4.1: Percentage of Phosphorus-32

In Figure 4.1 (Lawrence and Scott 1939, p. 696), Lawrence and his colleagues sacrificed their mice and measured radioactivity in different body regions per gram with electroscope and the Geiger-Müller counter. The Geiger-Müller counter is a tube with a high voltage applied to inert gas. Incoming radiation ionizes the inert gas and makes it conductive. The sudden change in conductivity produces an electric pulse that researchers can then detect. Using the Geiger-Müller counter and the electroscope, Lawrence and his team noted that, first of all, the total, full animal body radiation levels were similar between normal and lymphomatous mice. But lymphomatous mice
saw increased concentration of Phosphorus-32 in tumors but reduced Phosphorus-32 concentration in liver or bone, in comparison to normal mice. Researchers thought Phosphorus-32 concentrated in tumors “at the expense of bone and liver.” (Lawrence and Scott 1939, p. 695) Without tumor, Phosphorus-32 concentrated heavily in the bone. But with tumor, Phosphorus-32 concentrated at tumor sites. Lawrence and his colleagues summarized and hypothesized that perhaps Phosphorus-32 did not have any specific affinity for bone marrow. Instead, Phosphorus-32 merely concentrated at sites of heightened metabolism (p. 695):

It is clear that the total phosphorus content of a tissue does not necessarily determine the uptake or exchange of a given dose of phosphorus in that tissue. The latter is determined by the rate of metabolism of the element in question in a particular tissue. The higher activities in bone and tumor tissue suggest the use of radiophosphorus as a source of therapeutic irradiation in conditions involving primarily the bone marrow. Lawrence and his colleagues’ discovery was a critical insight into a potential connection between radiation and tumor detection. They demonstrated the existence of radioactive isotopes that selectively concentrate in tumor sites, and they even came up with a rudimentary hypothesis of the background mechanism. They unknowingly identified a category of radioactive isotopes now known as “perfusion tracers” – “perfusion” in the sense that the radioactive isotopes do not participate in biological functioning but just perfuse throughout the body; “tracer” in the sense that the radioactive isotopes have selectivity in where they concentrate. Lawrence and his colleagues’ work was interrupted by World War II (cyclotrons around the U.S. were drafted for military uses during World War II), but their early exploration enabled later generations to pick up the idea later. Their work was one of the very first cases (partially due to easy access to the new cyclotron at the time) to establish that some isotopes could selectively concentrate in tumors.
However, the work was also marred by significant shortcomings. For example, it had very low resolution. Lawrence and his colleagues were able to distinguish radiation from different body regions (the liver, the bone, etc.). They were able to tell that tumor sites had much higher radioactivity than, say, the liver. But those regions were way too crude to allow for precise localization. Not knowing specifically where the radiation came from prevented researchers from inferring whether radiation in the bone came from the bone marrow or not. Perhaps the shortcoming was a good thing: perhaps it is common among researchers to jump on a flawed but exciting new research direction, so they can find straightforward ways to contribute to a field that is not yet too crowded (for an example of such an effect, refer to Fujimura’s discussion of the early history of the proto-oncogene research: Fujimura 1997, p. 159). Not surprisingly, shortly after the war, other researchers joined them.

### 4.2.2 Incorporating Phosphorus-32 Into a Tumor Detection System

Despite the low resolution of their method and somewhat speculative hypothesis that Phosphorus-32 concentrated in sites of heightened metabolism, Lawrence and his colleagues’ insights intrigued researchers after the end of World War II. For example, in the 1950s, Sweet and his colleagues at MGH picked up the idea that Phosphorus-32 selectively concentrated inside tumors (Sweet 1951). Their interest was not to use radiation to treat tumors, but primarily to locate tumors.

The team at MGH was looking for practical solutions for a major challenge during brain tumor surgeries: tumor localization. Brain tumors do not look very distinct from normal brain tissue to the naked eye. Especially on the contours of brain tumors, identifying and separating brain tumors from normal tissue are extremely difficult. Researchers then and now are constantly looking for better ways to identify brain tumors and save normal tissue as possible. Sweet, the first author of the paper from MGH exploring Phosphorus-32 in brain tumor localization, picked up the discovery from Lawrence and his colleagues (Sweet 1951). They suggested that if Phosphorus-32 could selectively...
concentrate in brain tumors, and if researchers could detect Phosphorus-32 during surgery, they would have a way to remove only Phosphorus-32 rich brain tumors while sparing normal tissue.

The first step toward this goal was to find a better Phosphorus-32 detector. Phosphorus-32 could not penetrate the skull so its detection would require an opening of the skull and direct access to the surface of brain tissue. But, during surgery, surgeons only open up a very small portion of the skull. The detector must be small and nimble enough to pick up radiation from a small opening. But the Geiger-Müller counter used by Lawrence and his colleagues was simply way too big to dip through the small opening on the skull.

In order to localize brain tumors during surgery, researchers at MGH identified three desiderata (Robinson 1950, p. 82):

- Small physical size, so researchers could dip the Geiger-Müller counter through the small skull opening
- Good precision, so that the counter only responded to radioactivity from a very small region around the tip of the Geiger-Müller counter
- Fast counting speed, for quick results during surgery

The first desideratum, size, is in response to the specific pragmatic need in brain surgeries to fit into a very small skull opening.

The second desideratum reflects a general goal of higher resolution. I will return to this in the next chapter. Here I focus on researchers’ definition of “precision.” By “good precision,” they meant specifically responding to radiation only from a very small region:

For precise localization it is desirable that the sensitive volume be small and located near the tip of the counter (p. 82).
In the quote, researchers expressed a desire to record radioactive levels only from a very tight region near the tip of the counter. The goal was to ensure that whenever the new counter registered incoming radiation, researchers could be fairly certain that the incoming radiation came from somewhere within a very small region. In other words, the goal was to better separate incoming radiation from one small region to another small region and better infer to where the radiation came from. The concern ultimately is to ensure high quality inference, from a data point to the source of the radiation.

The third desideratum reflects goals of higher sensitivity and faster detection. The team at MGH noted that there normally is a trade-off between a smaller radioactive response region (higher spatial resolution) and a slower counting speed (Robinson 1950, p. 82). This is because the Geiger-Müller counter requires a certain minimum threshold of incoming radiation to register. If the counter takes in radiation from a small region, then it takes longer for it to accumulate enough radiation to pass the threshold. The researchers wanted their device to respond faster, but not primarily because they cared a lot about speed, or temporal resolution (a topic for the next chapter). After all brain tumors do not change significantly during the course of a surgery. The location of brain tumors identified an hour ago would work perfectly fine for surgical purposes an hour later. But they wanted the detection instrument to be more efficient: they did not want to keep counting at a small region for a long time during surgery, or wasting or ignoring information if radiation did not pass a certain threshold to be registered. In other words, the team was primarily interested in making more efficient uses of useful information and they were concerned with increasing sensitivity.

In short, researchers explicitly listed the constraining factors in their search space for a new detection instrument. In the list, we see one primarily pragmatic concern – reducing physical size, and two general normative concerns – increasing resolution and increasing sensitivity.

Researchers then went on to identify a good solution through trial-and-error. As we saw in the previous chapter, listing possible candidate materials and proceeding with trial-and-error with all or at least most of them to find a good solution was not a foreign idea to Kallmann, Bell, and
others. The MGH team first tried two common inert gas mixtures, ether-argon mixtures and ethyl acetate-argon. They concluded that, other things equal, ethyl acetate-argon produced a much longer and detectable pulse. They then systematically varied components of their design, as demonstrated in the following table:

<table>
<thead>
<tr>
<th>Counter No.</th>
<th>Cathode diam. mil</th>
<th>Wire diam. mil</th>
<th>Ethyl acetate pressure mm</th>
<th>Argon pressure atmospheres</th>
<th>Threshold volts</th>
<th>Plateau length volts</th>
<th>Plateau slope %/100v</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>228</td>
<td>1</td>
<td>25</td>
<td>2</td>
<td>1120</td>
<td>275</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>112</td>
<td>1</td>
<td>25</td>
<td>2</td>
<td>1025</td>
<td>150</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>67</td>
<td>1</td>
<td>15</td>
<td>1</td>
<td>710</td>
<td>70</td>
<td>17</td>
</tr>
<tr>
<td>4</td>
<td>67</td>
<td>1</td>
<td>25</td>
<td>1</td>
<td>850</td>
<td>80</td>
<td>13</td>
</tr>
<tr>
<td>5</td>
<td>67</td>
<td>1</td>
<td>25</td>
<td>2</td>
<td>910</td>
<td>100</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>67</td>
<td>1</td>
<td>30</td>
<td>3</td>
<td>1000</td>
<td>125</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>67</td>
<td>1</td>
<td>35</td>
<td>4</td>
<td>1065</td>
<td>110</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>67</td>
<td>0.4</td>
<td>15</td>
<td>1</td>
<td>670</td>
<td>60</td>
<td>16</td>
</tr>
<tr>
<td>9</td>
<td>67</td>
<td>0.4</td>
<td>25</td>
<td>1</td>
<td>780</td>
<td>100</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>67</td>
<td>0.4</td>
<td>25</td>
<td>2</td>
<td>830</td>
<td>100</td>
<td>3</td>
</tr>
<tr>
<td>11</td>
<td>67</td>
<td>0.4</td>
<td>30</td>
<td>3</td>
<td>890</td>
<td>150</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>67</td>
<td>0.4</td>
<td>35</td>
<td>4</td>
<td>960</td>
<td>175</td>
<td>2</td>
</tr>
<tr>
<td>13</td>
<td>30</td>
<td>1.5</td>
<td>30</td>
<td>8</td>
<td>870</td>
<td>30</td>
<td>1</td>
</tr>
</tbody>
</table>

Figure 4.2: Experiment Data

In Figure 4.2, researchers listed a total of thirteen different designs, with varying cathode and wire diameters, inert gas pressures, voltage levels, etc. Not mentioned in the table was that they also experimented on the materials used in their design. They used degreased stainless steel for cathodes in number 2 and number 13 and activated copper for all others. After systematically varying the components, researchers ended up with a modification of test design number 13, deeming that design number 13 offered the best balance between three desiderata.

Armed with the compact Geiger-Müller counter, researchers at MGH set out to validate their new instrument. First, they had to confirm that Phosphorus-32 selectively concentrated in tumor sites
through a series of steps. They first injected patients with Phosphorus-32 ahead of surgery. They then surgically removed the tissue. They analyzed the tissue to determine its radioactivity and whether it is cancerous. It turned out that tumors were 5.8 to 100 times more radioactive than normal tissue. They performed such comparisons on a wide range of brain tumors including glioblastoma, astroblastoma, astrocytoma, and atypical glioma and confirmed that Phosphorus-32 did selectively concentrate in brain tumors (Selverstone and Solomon 1948; Selverstone, Solomon, and Sweet 1949). They relied on postmortem specimen to further validate their measurement. They reported four cases in which patients died a few days after operation and they could directly compare the records of operation, tumor assessment based on radioactivity, and the actual distribution of tumors. It turned out that identification of tumors based on radiation level was fairly accurate (Selverstone and White 1951).

Now with preliminary validation, researchers at MGH started to deploy their new instrument to detect brain tumors during surgeries. Once they opened up the skull during operation, researchers could now move their compact Geiger-Müller counter across the surface of the brain. If they noticed a sudden jump in radioactivity, they took it as a sign they had crossed the borderline between normal and tumorous tissue (Selverstone, Solomon, and Sweet 1949). As of 1951, they performed a total of 144 such surgeries. Post-surgery analysis showed that in 108 cases, they correctly identified tumor based on recorded radioactivity (Selverstone and White 1951).

The invention and validation process at MGH is exceedingly similar to Kallmann’s invention processed covered in the previous chapter. The shared feature of both research teams is the little role of theory-testing played in driving scientific and technological development forward. Here, I do not mean that MGH team or Kallmann had no knowledge of the relevant theories of radioactivity, etc. I mean that theoretical considerations, such as confirming or testing specific theories, played almost no role in motivating and driving forward their research. In MGH team’s case, their concern was to develop a smaller, more sensitive, and higher-resolution detection device. The driving forces were,
in other words, the availability of materials, pragmatic concerns, and engineering norms. To fulfill this concern, they experimented on a number of materials, based on the availability of materials, equipment, and experimental methods. They came to the invention of a new, smaller detection instrument. We can identify the same story from Kallmann’s invention. Here is a figure from the previous chapter.

![Figure 4.3: Kallmann’s Experiment Data](image)

In Figure 4.3, similar to the MGH team, Kallmann was concerned with finding a better material to stop and capture radiation. He relied on the availability of materials at hand and equipment to identify the most suitable solution to him.

The adoption of Phosphorus-32 in brain surgeries is partially thanks to increased resolution. During Lawrence and his colleagues’ days at Berkeley, they could only identify Phosphorus-32 deposits in broadly defined regions such as “the liver” and “the bone.” They did not have better measurements of radiation, especially with high resolution, to infer where in the liver or where in the bone the radiation came from. With the improvement at MGH, researchers could now fairly accurately determine the contours of brain tumors. This is because the new equipment enabled them to better separate radiation from different regions (thanks to the fine tip of their new Geiger-Müller counter) and better infer the source of the radiation.
Such preliminary validation also led researchers at MGH to see many limitations of their new instrument. First, they established that their technique did not have high enough resolution to detect diffused tumors. It was good at detecting generally confined solid tumors. But it could not precisely pinpoint the contours if tumors were diffused and the line between tumors and normal tissue became blurrier. Second, they were aware that the beta ray emitted by Phosphorus-32 during its decay could only penetrate roughly 7.5 mm in brain tissue (Selverstone, Solomon, and Sweet 1949). Researchers could not rely on Phosphorus-32 to detect tumors without opening up the skull or localize tumors deep inside the brain. For example, tumors with over two centimeters inside the brain were not detected (Selverstone and White 1951).

New understanding of the limitations of the instrument gave researchers ideas for future research. First, they needed radioactive isotopes that emitted energy that could travel from deep within the brain. Further, if the radiation could penetrate the skull, they would not even need to open up the skull and could use radioactive detection to plan the surgery ahead of time. Third, if they wanted to use radiation that traveled far, they needed new detection instruments. Unlike beta rays emitted by Phosphorus-32, gamma rays travel very far and can penetrate bones (X-ray is a form of gamma ray). Gamma rays also do negligible harm to normal tissue, making gamma rays safer to use. However, the Geiger-Müller counter was designed to capture beta rays and was inefficient at detecting gamma rays (Selverstone and White 1951). They needed new instruments.

Trying to address the aforementioned challenges, researchers at MGH turned their attention to positron detection technologies developed by around the 1950s, as described in chapter 1.

4.2.3 Positron-Based Detection of Brain Tumors

In the previous chapter, I mentioned that Wrenn et al. were among the first teams to realize that a paired-scintillator design with coincidence detection could improve image quality. The scintillator consists of two components. It has a crystal, which “flashes” when hit by some radiation (hence the name, “scintillator”). And it has a photomultiplier, which collects and amplifies the weak flash. The
amplified flash can then be captured by conventional means, technologies not too different from those in our digital cameras. In other words, the scintillator first captures and converts radioactive energy into visible light, then the photomultiplier amplifies the light so that it can be better detected. Before the discovery of the paired-scintillator, coincidence detection design, researchers relied on a very low sensitivity design to screen off irrelevant information. Previous researchers were primarily relying on the collimator.

![Hooded Camera](Image)

**Figure 4.4: Hooded Camera**

The idea of a collimator is similar to the camera lens hood (as in Figure 4.4) on a quality standard photographic camera. Like the hood, the collimator restricts the incoming angle of radiation. Only radiation from predetermined angles can reach the crystal. Everything else will be screened off by the “lens hood.” The background theory for using collimator is that scattered radiation will come from all kinds of unpredictable directions. But unscattered radiation will come in at an angle perpendicular to the crystal. This idea is not fully incorrect: a collimator can screen off most scattered radiation. But it screens off most unscattered radiation as well, resulting in very low sensitivity of the instrument. The collimator design requires a lot of radiation to register hits and hence wastes a lot of radiation.

In order not to waste useful information, and raise the sensitivity of the instruments, researchers have been investigating ways to preclude scattered radiation without also precluding significant
amount of unscattered radiation. They were looking for precise, instead of blunt tools, to screen off scatter radiation. Wrenn et al. pointed out that paired-scintillator design did not rely on the collimator to filter out useless information that discarded useful information at the same time. They discovered that a key distinction between useful and useless radiation, after an annihilation event, was the time and location of radiation’s arrival. They hence designed a pair of scintillators that registered a “hit” only when two gamma rays arrived at the same time and in opposite directions. For this reason, the registration method is also called coincidence detection. If a pair of gamma rays did arrive at the same time, Wrenn et al. could infer that some annihilation event happened along the straight line between where those two gamma rays arrived, also called a line of response. Instead, if gamma rays were scattered, they would (barring random coincidence) not arrive at the same time or at roughly 180 degrees apart from each other. By using the paired-scintillator design, instead of the Geiger-Mueller counter, Wrenn et al. reached a level of signal-to-noise ratio, sensitivity, and precision way above the level of the Geiger-Mueller counter (Wrenn, Good, and Handler 1951). Researchers at MGH independently came up with the same solution and arrived at the same conclusion (Sweet 1951).

Using gamma rays, as mentioned above, has a few more advantages. Unlike beta rays, gamma rays can travel very far. Researchers do not need to open up the skull to pick up gamma rays from the outside. Gamma rays also do negligible harm to normal tissue, making gamma rays safer to use.

But the paired-scintillator design at the time was more a concept than a reality. Wrenn et al. only made some quick demonstration of the design’s advantage, without showing how one could apply the design in real applications to localize brain tumors (Wrenn, Good, and Handler 1951), partially because they had no idea of a suitable radioactive isotope. Researchers at MGH decided to try out the combination of some positron emitting radioisotopes and coincidence detection. They first tried a gamma-emitting isotope, K^{42}. Unfortunately, Potassium-42 did not work. Muscles seize Potassium-42 readily, causing safety worries (Sweet 1951). The short half-life (roughly 12 hours) also means
that the researchers have to order and ship Potassium-42 inconveniently often (Selverstone and White 1951). Researchers elsewhere at the time discovered and manufactured many radioactive isotopes for research in nuclear physics. Radioactive isotopes of interest to nuclear physicists were not necessarily safe to human beings or useful in detecting brain tumors. Sweet et al. at MGH pointed out that they needed a radioactive isotope that fulfilled the following criteria (Sweet 1951):

- The radioactive isotope should not be toxic to human beings

- The radioactive isotope should have a reasonable half-life – the time it takes to halve its radioactivity. Too short a half-life means that the radioactivity will be depleted before researchers can have an opportunity to experiment or finish surgery. Too long a half-life means that the isotope can keep releasing energy inside the human body for a long time, causing potential harm to the human body

- The radioactive isotope should selectively concentrate in brain tumors, so it would give out most radiation from within brain tumors, helping researchers localize brain tumors

The needed help came from neighboring disciplines not focused on using radiation for medical uses. Chemists and physicists realized the value of the paired-scintillator design. They could infer from the number of gamma rays detected to the number of positrons at the site of annihilation. Because the concentration of externally introduced positrons often depended on the internal structure of the solids and liquids under investigation, researchers elsewhere quickly started to use the paired-scintillator design to study the internal structure of solids and liquids (Wallace 1960). Chemists and physicists’ interests led to a growing list of radioactive isotopes developed and ready for Brownell’s team to choose from. Of course, other chemists and physicists mostly did not care about the biomedical properties of their radioactive isotopes. They were primarily interested in expanding the reservoir of radioactive isotopes for their own uses, although the expanded reservoir provided Brownell and his team with expanded selection as well. After around 3-4 years of research, Brownell and his team seemed to have the suitable isotope candidate. Their primary concern was the half-life of isotopes.
Carbon-11 had only 20 minutes of half-life, too short for their uses. Many other candidates had way too long a half-life to be safe. So Brownell and his team settled with Copper-64 initially (Brownell and Sweet 1953). Copper-64’s half-life is about 12 hours, similar to Potassium-42 and just as inconvenient to order and replenish, but it is safe for human muscle.

Armed with the new-found isotope, Brownell and his team started to design their scanners. Their design of the first prototype consisted primarily of a pair of scintillators facing each other. But their first design defeated the purpose of improving sensitivity. Improving sensitivity partially means to capture and utilize useful information as possible. But merely a pair of scintillators will miss most of the useful, unscattered radiation traveling in all 360 degrees.

To better capture escaping gamma rays and improve sensitivity, Brownell and his team designed an automatic mechanism to move their scintillators. They attached the pair of scintillators to a movable platform. The mechanism first moved the camera horizontally, so that the camera could slowly detect and record the escaping gamma rays on one horizontal plane. Once done with one plane, the mechanism shifted the camera vertically and reversed the horizontal movement to detect and record gamma rays on a different horizontal plane. The process went on until the scintillators finished scanning the whole brain (Brownell and Sweet 1953).

In Figure 4.5 is the prototype with its associated movement mechanism developed in 1953. A test subject was placed between the two cameras. The person on the left in a lab coat was Brownell, the leader of the MGH team. The person on the right is Aronow. Aronow had a physics PhD from Harvard University. He was both a faculty member at Harvard Medical School and the lead of equipment development at Brownell’s Physics Research Lab. The prototype was called the “[f]irst clinical positron imaging device” (Brownell 1999, p. 2).

But the first clinical positron imaging device worked only by chance. Brownell and his team first spent a lot of time on figuring out the parameters of the system. For example, they studied how sensitive the crystal was and how much noise there was in the system. Armed with this knowledge,
they experimented on human subjects. It turned out that the image was barely useful. This is visually demonstrated in Figure 4.6. In that figure, each “tick” or block was a spot where the camera system stopped and took a reading. The darkness of the reading was indicative of the intensity of the radioactivity reading. The team then superimposed an image of the brain onto the reading. Being among the first teams to perform full head scanning with a novel positron detection system, the MGH team did not have much knowledge or experience in developing a protocol of experiments or processing and visualizing the data. And the image came out poor and hard to explain. Merely looking at the image (Figure 4.6), it was incredibly hard to tell if there were brain tumors or where they were. However, when Brownell and his team moved the patient off-center, placing the patient closer to the left scintillator than to the right one, the image (Figure 4.7) all of a sudden became informative and useful. They had no idea why at the time.
Figure 4.6: The Result of a Balanced, Intended Scan

Figure 4.7: The Result of an Unbalanced Scan
4.2.4 From the First Prototype to the First Positron Camera

After the first prototype in the 1950s, research at MGH stalled for a whole decade. The lack of progress was partially attributable to a lack of marketable application in which the device could solve problems in ways others could not. First, the prototype largely failed its original purpose. It was designed to image brain tumors to guide the surgery. But the prototype, like the pictures above, offered only two-dimensional images. Two-dimensional images are of limited use during brain surgery, because precise operations require three-dimensional information. Even with the two-dimensional images, researchers struggle to interpret those images. For example, in Figure 4.7, interpretation of the image was nearly impossible without already knowing the answer. Consider a few “roadblocks” to the interpretation of the image:

- What do the “ticks” outside the face mean?
- What does the shadowy region around the face mean?
- What does the “white gap” between the shadowy region on top of the brain and the shadowy region around the face mean?

The diagnosis of “brain tumors” would make sense only if researchers already knew there were brain tumors in the darkest region and no tumors elsewhere. Otherwise, it would not be clear what kind of interpretation a researcher would make. Second, the prototype did not find appeal outside brain tumor surgery planning. Because the line-by-line imaging process was extremely slow, the prototype could be useful only in imaging phenomena that did not move and lasted a long time, such as brain tumors. It would not be useful in studies of, for example, stroke, blood circulation, etc. Lastly, although the prototype worked in practice, researchers lacked a theoretical understanding of why it worked when the patient was off-center. Fourth, as mentioned above even the two-dimensional images produced by the prototype were of limited use. The two-dimensional images were informative if we already knew what was going on.
Lastly, the prototype was incredibly inefficient and defeated the whole purpose of positron-based scanners. One advantage of positron-based scanners, as mentioned above, was their increased sensitivity. The Geiger-Müller counter needed to rely on the collimator to screen off most incoming beta rays to rule out the scattered ones. Positron-based scanners, based on coincidence detection, did not have to, because scattered gamma rays, by and large, would not arrive at the same time in pairs. Positron-based detection systems did not need to rule out most incoming gamma rays and hence should be more sensitive in collecting gamma rays. But the prototype completely defeated the purpose by having only one pair of scintillators. While positron annihilation results in a pair of gamma rays traveling in opposite directions, they can travel in any pair of opposite directions. One pair of scintillators could only detect incoming radiation in one pair of directions, hence wasting most gamma rays anyways. Not being able to capture most gamma rays partially contributed to the slow scan time and increased radioactive dosage to patients. And it contributed to very low sensitivity.

Ten years later, Brownell and his team tried to revive the first prototype to address the aforementioned issues. They wanted to improve the usefulness of the device for pre-surgery planning, increase sensitivity, and find a potential future for the instrument in fields outside brain surgery. They started with two research directions. The first effort was to construct three-dimensional (tomographic) scanners. Their idea of three-dimensional scanners was based on time-of-flight measurement. Previously, coincidence detection could tell only that an annihilation event took place along the line of response. But the “simultaneous arrival” was not exactly simultaneous, because insofar as the annihilation event did not take place precisely at the midpoint between the two detectors, there would be a small difference in travel distance and hence a small difference in travel time. Because light travels at roughly 300000000 m/s, a difference in 1 centimeters of travel distance translates into a difference of 30 picoseconds in travel time. If scanners were able to detect the small difference in travel time precisely, researchers could calculate exactly where the annihilation event took place along the line of response. Brownell and his team labored to increase temporal
resolution of their scanners. They managed to push down the temporal resolution significantly to the scale of nanoseconds (still not on the scale of picoseconds), but at a significant cost: they had to use radically inferior and inefficient crystal and scintillator design. For example, the otherwise ideal NaI(Tl) crystal had a response time of more than 200 nanoseconds, not suitable for fast detection. Brownell and his team had to make do with far inferior crystals. The computational circuits needed to be radically simplified too. In the end, Brownell and his team gave up on this attempt, leaving it to future generations of researchers with improved techniques (Brownell et al. 1969, p.166).

The second effort was a straightforward one. Instead of using merely one pair of scintillators, Brownell and his team decided to use nine pairs. The use of nine more pairs increased the sensitivity of their scanners because now they could capture gamma rays traveling in more directions (p. 170).

For a while, both directions – building three-dimensional scanners and increasing the number of scintillators in their system – seemed promising. On the three-dimensional images side, “improved techniques” came around quickly. Chesler, another member of the team, came up with a method to produce three-dimensional images out of two-dimensional scans. His method, known as the filtered back projection, was an engineering breakthrough. In a nutshell, their scanners offered only the information that an annihilation event took place along the line of response. Although researchers could not, based on merely one line of response, guess where the annihilation event took place, researchers could, based on a large number of lines of response, reverse engineer the received gamma rays and ask: what kind of distribution of annihilation events could produce the very specific pattern of received gamma rays as we had received them? Chesler’s reconstruction method turned out to work reasonably well with simple objects such as dummy bottles and other objects of various shapes (Chesler 1973).

At the same time, Brownell and his team kept on multiplying the number of scintillators on their scanners. In their next iteration of scanners, instead of nine pairs of detectors, they used a total of 127 pairs of NaI(Tl) crystals. In theory, each crystal should be paired with one photomultiplier. But
Figure 4.8: The Second Prototype, Called the Hybrid Scanner
more than a hundred photomultipliers would raise the cost beyond reason. To cut the cost, Brownell decided to pair one photomultiplier with more than one crystal and came up with the following ring design in Figure 4.9: Although each photomultiplier was shared by more than one crystal, any two photomultipliers could uniquely determine which crystal a flash came from (Brownell et al. 1969, p. 170). With the ring design and the large number of crystals, Brownell and his team came up with the scanner that they referred to as Positron Camera (PC).

Brownell and his team then went through another round of incremental improvement to produce the fully commercialized PC-II (Positron Camera II). The difference between PC-I and PC-II was limited. PC-II improved the motion design (the pattern of moving scintillators) so that with sufficient overlap of scanning the same spot from different angles, researchers could use post-processing to interpolate the pictures, increasing Positron Camera’s resolution by 30% (Brownell and Burnham 1973, pp. 158-162).

But both development directions ran into a dead end. Brownell and his team did not put Chesler’s reconstruction method into use. Nor did their new PC-II design fully replace PC-I with PC-II’s new theoretical advantage in improved resolution.
Both directions were more or less abandoned due to the conflict with a new, third direction of development. The third direction of development was partially due to the new cyclotron at MGH, which was installed in 1967, because of increased popularity of using radioisotopes in medicine (Tilbury et al. 1975, p. 521). Previously, Brownell and his team were using perfusion tracers, tracers that spread into tissue through blood flow, without participating in biological functions. Cyclotrons are able to generate short-lived isotopes such as 15-Oxygen, which participates in metabolism. In other words, while researchers could use perfusion tracers to detect the anatomical and blood flow structure inside the body (for example, tumors normally have abnormal blood flow structure), they could use metabolic tracers to detect metabolic changes - changes that were difficult to detect with other methods at the time.

Brownell and his team seized on this crucial opportunity and developed the “dynamic mode” for their scanners. The dynamic mode, in short, sacrificed spatial resolution and sensitivity for temporal resolution. It also changed the format of presentation. Because the human eyes could not readily detect nuanced changes across images, Brownell and his team decided to use charts, instead of direct depiction of organs, to demonstrate dynamic change (Brownell et al. 1971, p. 166).
The choice of presentation here caused long-term challenges in the history of the development of PET. The first challenge was about demonstrating nuanced change. Whether researchers should illustrate changes as numbers and charts (as in Figure 4.11) or as pictures (as in Figure 4.6) became a big debate. Second, because researchers opted for faster scanning at the cost of lower resolution and lower picture quality, they created long-term impact that protocols to set up experiments and to process data have to be designed with the assumption of low picture quality and with the end goal of compensating for it. One lasting impact is the focus on population average over individual differences in neuroimaging, partially because population average, for a long time, was much easier to extract and calculate with low picture quality and many research protocols and directions were developed to work solely with population average.

But for now, the immediate result was the abandonment of other directions as the team turned their attention to the dynamic mode. PC-II never fully replaced PC-I and actually its resolution advantage was never widely used, because Brownell and his team never found traction (Brownell and Burnham
1973, p. 164). The sacrificed resolution and sensitivity in the dynamic mode rendered the theoretical reconstruction suggested by Chesler’s method statistically impossible (Brownell and Burnham 1973, p. 157).

Although the MGH team paid a lot of attention to the third direction, the new direction never caught on either. While it was nice to understand the rate of Oxygen being cleared out of the pulmonary artery, it was never too clear what kind of new question the new data could answer. Especially given the million-dollar price tag of the cyclotron, it was not clear at the time if the investment would generate enough new answers that other, much cheaper instruments could not offer. The research at the MGH slowly died down partially because it never found a market: a market in which there are new questions that can be answered, and preferably only, by their Positron Camera. It would be another 20 years before researchers found a marketable use for positron emission technologies in the medical world.

In short, the history of Positron Camera, through the 1950s, is largely a technology-driven history. The MGH team had limited theoretical understanding of brain tumors or how Positron Camera measured brain tumors. Or else, they would not be puzzled by the better image produced by the off-centered patient. Their technological development is primarily a bottom-up and opportunistic history. They relied on and built on the availability of a nearby cyclotron, recent chemical developments in new radioactive isotopes, new crystals, and other technological developments. Out of those opportunities present in the scientific community, the MGH team improved Positron Camera along a number of engineering norms, such as increasing signal-to-noise ratio, increasing sensitivity, and increasing resolution, while looking for popular applications in medical research and for clinical purposes.
4.3 Toward A Philosophy of Sensitivity

In the previous historical section, I have reviewed how the team at MGH had developed a series of radiation imaging technologies. It started with a very accidental discovery at Berkeley before World War II: some isotopes, later known as “perfusion tracers,” selectively concentrated at sites of heightened metabolism. The MGH team picked up Lawrence and his colleagues’ insights and turned them into a tumor detector: tumor sites are often sites of heightened metabolism. The team subsequently developed a series of prototype and commercial detection devices, for a wide range of purposes, from pre-surgery planning to studying oxygen clearance from organs, driven by a desire to find new and popular applications for their new device and by the availability of new technology, such as the cyclotron. While the practical considerations and directions of their research were constantly changing, the MGH team was sticking to a stable set of normative values.

In particular, in this chapter, I focus on their interest in improving the sensitivity of their devices, no matter what the devices were designed for. In the move from the Geiger-Müller counter to the first Positron-based Prototype, moving from detection during surgery to detection for pre-surgery planning the MGH team increased the sensitivity of their detection systems. In the move from the Positron-based Prototype to the Positron Camera, the MGH team again increased the sensitivity of their detection systems. What exactly is sensitivity and why does it matter? I devote this subsection to its definition and significance.

Sensitivity, also known as Efficiency or Detection Efficiency, is a measurement of detection systems. It measures the percentage of useful information in the available information pool that is actually captured and identified as useful information. It is focused on the twin critical scientific and engineering challenges of capturing useful information and identifying useful information as useful. The challenge of capturing and identifying useful information may be familiar to anyone with civic engagement experience. Assume, for a moment, we are still using handwritten ballots. To count
handwritten ballots, electoral officials have at least two challenges. First, they have to make sure that voters actually fill out the ballot in the right way: they use crosses to denote their favored candidates, or write clearly if they decide to write in a name, or they sign in a consistent way. Second, they need to make sure that manual processors actually recognize the valid ballots as valid: all the content is readable and conforming and the signature is correctly identified. The challenge of doing both during an election is often tremendous for any large-scale election. And the same kind of challenge is present in the design process of detection instruments. It is an epistemic struggle to evaluate and assign meaning to information. It is a further scientific and engineering struggle to make machines do it. In the case of history of PET, the challenge is first to capture radiation – which is normally not seen or felt by human bodies, and then identify unscattered gamma rays as unscattered gamma rays.

Like signal-to-noise ratio, sensitivity is very much a distinct engineering measurement. It is distinct from epistemic virtues such as accuracy in very much the same way as signal-to-noise ratio:

- Sensitivity is a measurement of instruments, instead of data
- Sensitivity is diagnostic and offers concrete directions to causally improve instruments
- Sensitivity is about specific inference challenges in science – inference from data to phenomena

It is a measurement and derivatively a guideline for designing causal relationship to interact with information in the world. In this norm, the causal mechanism (detection instruments) is not “transparent” or taken for granted. Instead, it is the center of the entire enterprise: sensitivity is a common measurement on the specification sheet of a detection instrument and, as I am to cover later, offers detailed ways for researchers to improve the sensitivity of their instruments. Further, improving sensitivity will address a specific inference problem – identifying useful information as useful.

The history of the MGH team also demonstrates the ubiquitous presence of increasing sensitivity as a general epistemic and engineering norm in the development of detection instruments. The MGH team was struggling to find directions and futures for their machine. The practical goals kept
changing and priorities kept shifting. At a time, they were pursuing three directions at the same time: developing tomography, multiplying the number of detectors, and adapting to a new class of isotopes. Despite those changes, the MGH team always had increasing sensitivity in mind. The norm of increasing sensitivity is an invariant concern to MGH and many other teams. To better explain the scientific background of sensitivity, I rely on classic engineering design textbooks. Many engineering design textbooks and professional literature often cover sensitivity in a dedicated section to both explain the notion and give students and professionals advice on best design practices. The goal of this chapter is not to repeat best design practice advice. The goal is to dissect “best design practices” and figure out why those practices are considered good and shared by professionals. The goal, in other words, is to discover the assumed but often unspoken epistemology of experiments and instruments from mathematical formulas and pieces of advice. I quote primarily from Physics in Nuclear Medicine, partially because it is a classic in the field, now on its fourth edition. In this and in other pieces of literature, for uniformity, I use “sensitivity” to refer to terms of “Efficiency” or “Detection Efficiency.”

Physics in Nuclear Medicine defines “sensitivity” as follows (Cherry, James A. Sorenson, and Michael E. Phelps 2012, p. 155):

[Sensitivity] refers to the efficiency with which a radiation-measuring instrument converts emissions from the radiation source into useful signals from the detector.

Because efficiency can refer to many ways in which instruments or other systems use information, the authors here focus on detection efficiency. This is the idea I attempt to capture with my notion of sensitivity. Sensitivity is a measurement of how much of the available useful information is actually captured and processed by the detection system as useful information relative to problems of interest to researchers. It is often expressed as a percentage, because a conceptual formula for sensitivity is:
Sensitivity = 
\[
\frac{\text{Information Detected, Processed, and Identified as Useful Information}}{\text{Total Available Useful Information}}
\]

It is a measurement of how much the detection system utilizes the available pool of useful information. If a detection system has a high sensitivity, it can capture and process a lot of useful information from a small pool of useful information. If a detection system has a low sensitivity, it captures and processes only a small fraction of the available pool of useful information.

Increasing sensitivity is a common, general norm in detection system design, as noted in *Physics in Nuclear Medicine* (Cherry, James A. Sorenson, and Michael E. Phelps 2012, p. 155):

In general, it is desirable to have as large a detection efficiency as possible, so that a maximum counting rate can be obtained from a minimum amount of activity.

In a way, we can understand the desire to increase sensitivity through the analogy with elections I have used before. In all elections, to make sure that every vote counts, we want to record every legitimate vote as possible. The transition from handwritten to machine-readable ballots is perhaps one such effort to ensure more valid ballots are recognized as so.

The considerations that drive researchers to improve sensitivity in their instruments is directly analogous to the considerations that drive our electoral officials. The world offers a limited amount of useful information at any given time and we need to make best use of it – just like we need to make every effort to value every single valid ballot. If researchers can increase the sensitivity of their detection systems and extract more useful information, then they can:

- Gather more information in a shorter amount of time
- Gather enough useful information from a small pool of information
- Gather a potentially more comprehensive set of information
Let me explain each point separately. First, high sensitivity often directly translates into speed. If a detection system can capture more useful information from what is available, researchers can speed up the information collection process. Speed is important in a variety of ways. Because I am describing speed in the next chapter, I will use only some brief examples. For one epistemic example, many phenomena do not stand still. Faster imaging speed often directly translates to usable information. If one wants to take a picture of a racing car, but the picture-taking process takes more than a few seconds, then the resulting picture will be a blurry car running across the picture. For one practical example, time is a valuable practical resource. If a hospital can reduce the imaging time for each patient from 20 minutes to 10 minutes, then they can service more patients in the same time frame or motivate more patients to take the imaging service. The epistemic and practical considerations may interact in interesting ways too. In the early years of PET, because the imaging time took regularly half an hour or even an hour, very few test subjects were interested in donating their time for researchers to calibrate and improve their nascent technology. In the end, often researchers themselves had to double as test subjects to move the research forward.

Second, high sensitivity directly enables crossing preset thresholds despite a small pool of information. Very often, a detection system or some information processing unit requires a minimal amount of information in order to “register,” to “activate,” or to otherwise process information: our eyes require a certain level of luminance to see colors – below the threshold, everything is dark to us. For smaller pools of useful information, the pool can be so small that crossing the preset thresholds requires a very high percentage of this pool of information being collected and utilized in order to meet the absolute minimal amount of information needed. For example, still using the electoral example, candidates often need to gather a certain number of signatures to qualify for the ballot. Let us say that each candidate needs to gather at least 100 valid signatures to qualify. One candidate manages to put together 105 valid signatures, but the electoral system has only a 90% “sensitivity” – out of 105 valid signatures, it can only recognize 95 of them as valid. In such a system, the candidate will not be able to qualify although, of course, he or she should. But with 95%
“sensitivity,” the same candidate will qualify. Conveniently, sensitivity is precisely the measurement of the percentage of available, useful information being actually collected and utilized. In other words, whether the sensitivity is high enough often determines not to what extent, but whether a detection system can actually process the incoming information.

Lastly, high sensitivity can translate into comprehensive sampling of information. Let us return to the electoral example. Assume we have 10,000,000 valid ballots but the sensitivity of the electoral system is very low – only 1,000,000 votes are recognized as valid. Then we have reason to wonder if the electoral results are skewed, biased, or, in some other ways not representative of the entire pool of valid votes. However, if we recognize 9,998,000 valid ballots, then it is reasonable to assume that, despite a few misses, we have a good sense of voters’ desires and wishes. Higher sensitivity allows a detection system to get more and potentially better information out of the available pool of useful information.

For the MGH team, they wanted sensitivity for all of those reasons. They wanted a high sensitivity detection system to speed up the detection time – so as not to hold the Geiger-Müller detection in the same area for a long time during brain surgery. They also wanted a high sensitivity detection system to cross the Geiger-Müller detection threshold. Lastly, they wanted a high sensitivity detection system to be able to detect small, minor tumor tissue – which may not emit a lot of radiation and can be missed by a low sensitivity system.

The MGH team and other engineers also have systematic, mathematical guidance on ways to improve the sensitivity of their system. Sensitivity (S) is the product of four main factors (Cherry, James A. Sorenson, and Michael E. Phelps 2012, p. 156):

\[ S = g \times \epsilon \times f \times F \]

where \( g \) is Geometric Efficiency, \( \epsilon \) is Intrinsic Efficiency, \( f \) is fraction of output that can be picked up, and \( F \) is absorption and scattering.
Each of these warrants further exploration, as doing so will give us deeper insight into how it exactly measures detection instruments and offers guidance for future improvement.

### 4.3.1 Geometric Efficiency

Geometric Efficiency measures the loss of intensity of radiation as it travels. The form of radiation of concern here is *isotopic*, which means that radiation leaves the source in all directions equally. The further the radiation travels, the less intensity it retains. Physically and mathematically, the intensity of radiation is proportional to one over the square of the radius \(1/r^2\) from its source. In other words, if the detection instrument is scanning a small \(r\), such as the brain, it will capture radiation with a higher amount of intensity. If it is scanning a big \(r\), such as the whole body, it will capture radiation with a lower amount of intensity, most of which is “wasted” on the way out to reach the detection instrument. To ensure better sensitivity, the MGH team and later teams focused on the brain (instead of, say, the entire human body) and especially small focal brain tumors, because of the small \(r\) of those objects.

The role of Geometric Efficiency becomes more important in the next chapter. During the time between history covered in this chapter and history cover in the next chapter, researchers designed Positron Emission Tomography as a general-purpose machine: it will become big enough to scan animals, the whole body, etc. The immediate trade-off for a general-purpose machine is its loss of sensitivity due to Geometric Efficiency. The machine becomes some kind of “jack of all trades, but master of none”: it is flexible enough to do many kinds of imaging and detection work, but not really of high sensitivity to do many well. In the end, as to be covered in the next chapter, researchers at Washington University in St. Louis decided to refocus on the brain, shrink the size of their device, and increase sensitivity.
4.3.2 Intrinsic Efficiency

Intrinsic Efficiency refers to all the causal, intrinsic aspects of a detection system that influence sensitivity. The list of potentially interfering aspects is long and under-defined, partially reflecting the fact that, even in a machine completely designed and put together by human hands, our understanding of the internal causal structure is limited.

Consider a few examples, in conjunction with the history of Positron Emission Tomography. There are a variety of reasons why a detection system might not be able to efficiently capture incoming radiation. One factor is the choice of crystal: different crystals respond to different radiation differently. Readers might recall from Chapter 3 that some crystals are better at capturing beta rays and some are better at capturing gamma rays. Some can capture more and others can capture less. For this reason, Kallmann had to experiment on a significant number of crystals, figure out their physical properties, and integrate their physical benefits and downsides into his design. The design question does not merely stop at the choice of crystals. Different crystals require different lengths to stop radiation. If they are not thick enough, highly energized radiation may be able to penetrate and pass through (think of how X-ray can pass through bones) without being captured and recorded. So researchers have to balance the requisite thickness of the crystal with other design desiderata, such as space limitation.

Another factor is the placement of detectors. Radiation is isotopic and traveling in all directions equally. It means that if researchers design the detection system to be merely a stationary pair of detectors, then they can capture only a very small portion of radiation. Making the pair of detectors moving (in the case of the team at MGH, through a slow-moving platform) can increase the sensitivity of the system because it can capture radiation from more angles. Having more pairs of detectors and covering more angles, as the team at MGH eventually moved to a 127-pair design, can also increase the sensitivity of the detection system. Eventually, other competing teams, including
the team at Washington University in St. Louis to be covered in the next Chapter, developed new designs that covered the entire 360 degrees.

4.3.3 Fraction of Output that Can be Picked Up

Fraction of Output that Can be Picked Up is another general term to describe the percentage of information from the output that can be picked up. I have mentioned that sensitivity is a measurement of the percentage of the available useful information picked up as useful information. In terms of designing the detection system, it means that researchers have to both pick up the use information and recognize the information as useful. Fraction of Output that Can be Picked Up refers to the second step.

Let me use the mechanism of coincidence detection to explain this concept. Coincidence detection, in our radiation detection context, means that if two opposite detectors detect a simultaneous arrival of two gamma rays, the detection instrument would register a “hit.” But how does that happen? In Figure 4.12 (Cherry, James A. Sorensen, and Michael E. Phelps 2012, p. 121), we can see two signals arriving at the opposite detectors. If they arrive too far apart, once combined, they will form two “peaks,” and the system will not register anything. If they arrive close to each other, once combined, they will form one “peak,” and the system will record a “hit.” However, where should we place the threshold (the dotted line in the picture)? If researchers place the threshold too high, they are going to miss many genuine “hits” and reduce the general sensitivity of the system. If they place the threshold too low, they are going to include false “hits,” or arrival of gamma rays that are too far apart from each other and increase the false positive rate. There is no “natural” or obvious choice in situations like this and other aspects of the detection system (such as precision of pulse detection) will place limitation on the choice of the threshold. In any way, any choice of threshold means that researchers either miss at least some true positives or admit some false positives.

In the coincidence detection scenario, the challenge was about finding the right balance between false positive and true positives. It is merely one example of the challenges in recognizing useful
Absorption and Scattering

Absorption and Scattering measure the extent to which interactions of radiation outside the detector may reduce the amount of and information of radiation eventually available to the detection instrument. For example, the team at MGH’s first major decision was to address absorption and scattering. They initially used beta rays to detect tumors, but analysis quickly showed that beta rays cannot penetrate more than two centimeters of brain tissue. They are absorbed on their way out (and cause much damage to brain tissue in the process). For this reason, the team at MGH decided to go with gamma rays, which could penetrate thick brain tissue. But the team at MGH faced a second challenge: brain tissue would scatter gamma rays. Scattered gamma rays, besides extremely difficult reconstruction, would not carry information about the location of their source. To address
this challenge, the team at MGH developed coincidence detection, which greatly improved the sensitivity of their system.

Similar to signal-to-noise ratio, sensitivity is a measurement built based on causal understanding and on specific guidance for improvement. It is not merely an evaluation of instruments. The measurement, as a mathematical formula, integrates preliminary directions to understand why a machine has the sensitivity that it does and how researchers can improve the machine. The measurement of sensitivity covers causal relationships outside detection instruments: the absorption and scattering process and the natural loss of intensity during transmission. It also covers causal relationships within detection instruments: how the incoming information is both captured and recognized as useful information. In the end, sensitivity is an unapologetically causal measurement: it is built deeply on the assumption that the above mentioned relationships, about how radiation travels through tissue and the space and how a human-made device can detection radiation, are real and causal.

4.4 Conclusion

In this chapter, I focused primarily on the development of Positron Camera at Massachusetts General Hospital in the 1950s to tell a story of technology-driven scientific change and the value of sensitivity in the process. The history of Positron Camera is a history of searching: the team at MGH tried their best to look for applications for their nascent technology and they could not say that they were entirely successful. The technology was new, difficult to operate, and quite costly, while it was not immediately clear at the time, what kind of new questions could be answered by going through the hassle. It would be another twenty years, and with many iterations and improvements, when researchers found widespread adoption of a positron- and coincidence detection-based imaging technology. In the entirety of the process, the primary emphasis on the technology had been, instead of on verifying scientific theories, collecting scientific evidence, the availability of existing tools and modules and how a research team could adopt and re-arrange those modules to find new scientific
uses. It is a case in which the technological concerns, instead of scientific, theoretical ones, were in the driving seat of the research project at MGH.

In the process of technology-driven scientific change, sensitivity and increasing sensitivity, along with other engineering measurements and their associated norms, played an imperative role in shaping where the development goes. Similar to signal-to-noise ratio, sensitivity is a unique measurement in at least three primary ways. First of all, sensitivity is instrument-centric. Sensitivity is by itself a measurement of detection instruments. It is determined by causal experiments: feeding detection instruments with known information and evaluating how much of the input have been captured and identified correctly in the output. The measurement is further broken down into causal elements, such as the distance that radiation has to travel and its loss of intensity due to the distance. The account of sensitivity feeds into a general account of instrument-centric perspective that includes:

- Researchers can evaluate the quality of detection instruments by evaluating elements of their causal mechanisms according to the relevant engineering and detection norms.

- The assessment of the epistemic value and quality of detection instruments anchors the assessment of the epistemic value of data generated by those detection instruments.

Second, sensitivity is concrete and diagnostic. It informs researchers of specific ways to improve sensitivity. For example, the measurement of sensitivity accounts for absorption and scattering of radiation. An obvious direction for improvement is to reduce absorption and scattering. Indeed, in the history of Positron Camera, researchers replaced beta ray detector with gamma ray detector to reduce absorption and used coincidence detection to address the issue of scattering.

Lastly, sensitivity is not a measurement *in virtue* of testing theories or performing experiments. Sensitivity is designed to provide a specific angle for the general challenge of inferring from data to
phenomena: how much of the useful information from the phenomena is captured and identified as useful information. This is a critical input for philosophers to make such an inference.

The significance of instrument-centric measurements and norms partially rests on the philosophical discussion of experimenters’ regress. Experimenters’ regress, according to Collins, is the circular definition that a good outcome is obtained with good instruments, but good instruments give out good outcome (Collins 1985, pp. 79-128). Franklin suggests that the regress does not exist because researchers can always demonstrate that instruments are good by calibrating them – very often by feeding them with known information, adjusting instrumental settings, and verifying if researchers can obtain the expected results from instruments (A. Franklin 1998, p. 34). While I am very much in agreement with Franklin, I want to push forward. I have demonstrated that what is “good” for data and what is “good” for instruments may well be two distinct categories. Instrument-centric measurements and their associated norms, such as increasing sensitivity, are not necessarily norms in service of or in virtue of their roles in experiments or theoretical testing. They are independent, unique norms about what makes instruments good, as instruments.

The next chapter, where I will cover the history of PET in the 1970s and 1980s and discuss the norm of increasing resolution, will serve as another experiment for the significance of instrument-centric measurements and their associated norms.
Chapter 5


5.1 Introduction

In 1977, two decades after researchers at Massachusetts General Hospital (MGH) created their prototype medical scanners and three years after the creation of the most recent and advanced scanner (known as Positron Emission Transaxial Tomography III or PETT III) based on similar principles to MGH’s scanner at the Washington University in St. Louis (WUSTL), WUSTL researchers were worried if their Positron Emission Tomography (PET) research program would survive.

The assessment of PET by the scientific community was dire. According to Web of Science, only three articles were published in 1977 using the most advanced PET scanners from WUSTL. Five in 1978. Four in 1979. The number of publications did not bode well for the new scanner. 1977 is also the year when a number of researchers published articles to persuade the academic and the clinical
communities that this nascent tool was not of significant value (For example, Budinger et al. 1977; Keyes Jr et al. 1977). Budinger and his colleagues, a major research team at Berkeley working on a competing medical scanner technology, noted that PET had low sensitivity, even at a fairly high energy level, exactly the issue the MGH team struggled with in the 1950s. Although PET could better utilize the data received, it required camera or patient rotation, which was challenging at the time (Budinger et al. 1977). They concluded, basically, that PET brought in very limited new or more information, at the cost of a million-dollar price tag to build not only the scanner, but also the accompanying cyclotron to manufacture the requisite, short half-life isotopes. The criticism was so overwhelming that a core member of the research team at WUSTL, Michael Phelps, had to publish an article titled "What is the Purpose of Emission Computed Tomography in Nuclear Medicine?" (Phelps 1977)

But by the mid 1980s, the tide has turned. In this chapter, I tell the turnaround story. The history between 1974-1981 is the history of iterative but somewhat uncontrolled improvement. In each iteration, the WUSTL team modified the PETT design to explore potentially new research directions and to find academic adoption of the PETT. At the same time, given that the research direction was constantly changing while the team was looking for academic interest, they kept improving PETT along a few normative criteria so PETT could become a good detection instrument in general. The history also teaches two philosophical lessons. First, the ultimate normative direction that the WUSTL team settled down with deserves philosophical attention. The WUSTL’s latest design of PETT VI was popular partially because it was fast and could image normal biological processes in real time and in vivo. The associated norm, increasing temporal resolution is an engineering norm discussed by Hacking but could receive more philosophical clarification in its concept and in its engineering significance. Second, the mechanism through which the WUSTL team identified and prioritized the norm of increasing temporal resolution also deserves philosophical attention. As I have iterated in this dissertation, most of the time, researchers face technical trade-offs between fulfilling different engineering norms. The history of WUSTL informs us that researchers and their
detection instruments could focus on a neglected norm, genuinely excel based on the evaluation of the neglected norm, and create a competitive landscape surrounding the neglected norm with few contenders.

### 5.2 History

In the previous chapter, I have noted when researchers at MGH were developing the very first prototypes of the PET scanner, they wanted the PET scanner to primarily detect tumors. Then the MGH team pivoted to include efforts to detect cardiovascular events, assessing the functionality of organs, etc. Tumor and cardiovascular diagnosis has always been a hotbed for new imaging techniques. Tumors do not always cause visible deformation on the outside and many of them look like normal soft tissue to the naked eye. Researchers hence need ways to identify and localize tumors before surgery so that they can remove tumors while preserving as much normal tissue as possible. This was the MGH team’s primary concern, as recounted in the previous chapter. Similarly, cardiovascular events often have no visible external signs and require ways to peek inside the body without unnecessary harm done to the body, such as through an invasive procedure.

To use PET scanners to detect tumors or cardiovascular events, researchers in the 50s and 60s depended on two kinds of positron-emitting isotopes. First, they could pair up the Positron Camera with perfusion tracers – radioisotopes that just diffuse into tissue but do not participate in or influence any major biological function. Some radioisotopes have selective affinity for tumors when they diffuse into tumors. With those radioisotopes, Positron Camera will pick up a heightened amount of radioactivity from tumor sites. Similarly, cardiovascular events are often coupled with changes in blood flow (obstruction, diversion, etc.). Perfusion tracers can illuminate exactly where blood flow is enhanced or reduced.

Second, researchers can pair up the Positron Camera with metabolic tracers – radioisotopes that participate in metabolism. For example, glucose is an indispensable metabolic participant. In order to
track glucose, scientists can “tag” glucose with Fluorine-18, which is a radioactive isotope of Fluorine. Fluorine-18-tagged glucose participates in metabolism, is “trapped” at the site of metabolism instead of being further processed due to biochemical differences from the non-radioactive glucose, and keeps emitting energy from the site of metabolism. Because tumors normally have higher metabolic levels, Positron Camera will pick up a heightened amount of radioactivity, this time out of Fluorine-18-tagged glucose, from tumor sites. For cardiovascular events, again, blood flow changes often lead to metabolic changes: low blood flow often means low metabolism in the relevant region. Metabolic tracers could demonstrate where blood flow changes and suggest cardiovascular events.

In 1974, when researchers at WUSTL created their first prototype of the PET scanner (named Positron Emission Transaxial Tomograph, or PETT), came up with a procedure to use PETT to detect tumors and cardiovascular events, and finished their first human scan, they found themselves at a significant disadvantage: their PETT was expensive. In order to image metabolic changes, their PETT requires radioactive isotopes such as Carbon-11, Nitrogen-13, Oxygen-15, or Fluorine-18. With the exception of Fluorine-18, the other three have very short half-lives. Carbon-11 has a half-life of 20 minutes, Nitrogen-13 10 minutes, and Oxygen-15 2.1 minutes (Wolbarst 1993, p. 447). Researchers at WUSTL primarily worked with Carbon-11 and Oxygen-15 because they had a cyclotron – a nuclear accelerator that manufactured those radioactive isotopes – on-site to produce those isotopes on demand. At the same time, short half-life isotopes are conducive to human experiments. Short half-life means a radioactive isotope will decay and become depleted in a very short time. If researchers use Oxygen-15, they only need to wait for a few minutes to run the next round of experiments on the same group of patients. It also helps that patients suffer from shorter radioactivity exposure.

But short half-life also means hefty costs. If researchers want to use those short half-life radioactive isotopes, they need to manufacture them on site. Manufacturing radioactive isotopes was and still is by no means cheap. WUSTL was one of the very few universities to have a cyclotron on-site in the
70s. In 1940, when WUSTL started to construct its first cyclotron, the cyclotron cost the school 106,000 USD, roughly equivalent to 1.5 million USD today after adjusting for inflation (Sherwood Moore, n.d.; Elliot Washington 1942). The installation took sixteen and a half months and the speed was considered miraculous. The installation of similar cyclotrons took Harvard 30 months and MIT 22 months (Elliot Washington 1942). The maintenance of the cyclotron was not cheap either. WUSTL’s first cyclotron’s rated Maximum Power was 72kw (The Washington University Cyclotron Dimensions and Characteristics 1955). Counting in all the necessary auxiliary equipment, the total normal power input to the cyclotron facility at WUSTL was 110kw, which is roughly equivalent to the normal electricity usage of more than 70 apartment units. Even the Medical School at WUSTL, which proposed the idea of an on-site cyclotron, found the cyclotron’s operational cost too high, decided to source radioactive isotopes from other manufacturers, and transferred the facility to the Department of Physics in 1946. The second cyclotron, installed at the Medical School, cost more than 500,000 USD in 1969, roughly equivalent to 4 million USD today (Mallinckrodt Institute of Radiology 1969). The second cyclotron was smaller, more efficient, and cheaper to operate. But the upfront cost was prohibitive to most other research universities or hospitals, to say the least.

By 1974, cheaper and commercially mature medical imaging techniques had been available on the market and a part of the academic and clinical routine for decades. The two major competitors to PET were X-ray Computed Tomography (X-ray CT) and Single Photon Emission Computed Tomography (SPECT).

### 5.2.1 X-ray CT

The X-ray CT scanner is the three-dimensional version of the well-known X-ray scanner. The X-ray scanner generates two-dimensional images of the internal structure of objects by passing X-rays through objects and evaluating the extent of attenuation of X-rays on their way out. Röntgen invented the X-ray scanner in 1895 (Röntgen 1896). G. N. Hounsfield invented the X-ray CT scanner in 1972.
The X-ray CT scanner witnessed swift academic and industrial adoption. Although the X-ray CT scanner was only invented two years before the PETT at WUSTL, researchers already had more than seven decades of experience in working with the X-ray scanner and had already deeply integrated the X-ray scanner into their academic and clinical routine. Further, Electric and Musical Industries (EMI), G. N. Hounsfield’s employer, helped the commercialization and adoption of the X-ray CT scanner. EMI, blessed by the Beatles fortunes, threw itself behind the X-ray CT scanner (Bates et al. 2012, pp. 105-108). In the same year as Hounsfield’s announcement, EMI began to manufacture and sell the X-ray CT scanner. They first manufactured twenty scanners right in Hounsfield’s research lab. In 1973, demand grew so fast that EMI moved to a dedicated manufacturing facility (p. 111). By around 1975, Ter-Pogossian, head of the PETT research team at WUSTL, estimated that around 600 units of X-ray CT scanners had been sold, taking up roughly half of the U.S. market for medical imaging equipment (M. M. Ter-Pogossian 1976, p. 1).

Coupled with the maturity of the instrument was the maturity of the corresponding research method. Researchers had long come up with ways to use the X-ray scanner to detect tumors (Fass 2008). Although tumors and normal tissue look roughly the same under X-ray, researchers found ways to infer the existence of tumors from structural anomalies. For example, researchers could use the X-ray scanner to detect bone cancer, because bone cancer often causes bone fractures or deformations. For another example, researchers could use the X-ray scanner to detect abnormal soft tissue contours. Also, researchers could inject dyes that selectively concentrate in tumors, which cause tumors to look differently under X-ray. However, for the most part, X-ray imaging of tumors was useful, but not definitive at the time. Cancer is only one of many potential causes for structural anomalies. Bone deformations can be the result of impact or ill-advised exercise routines. Researchers commonly need other means to confirm their inference from X-ray images.
5.2.2 SPECT

SPECT is a nuclear medical imaging technique similar to PET in basic principles. Unlike PET, which relies on paired gamma rays coming from the annihilation of positrons, SPECT relies on gamma rays emitted by gamma-emitting isotopes. After researchers inject radioisotopes into a living organism, the radioisotopes will decay and emit gamma rays. The SPECT scanner then picks up radioactivity from the outside and produces an image of radioactive concentration inside the organism. The technique is called “Computed Tomography,” because it uses the same basic principle as X-ray CT to acquire three-dimensional images. It is called “Single Photon Emission” CT, because the radioisotopes used for SPECT emit (hence “emission”) gamma rays (hence “photon”) (instead of positrons in the case of PET) when they decay. Researchers working with SPECT use diffusion tracers such as $^{99m}$Te that selectively concentrate in tumors.

While the development of the SPECT scanner was only a few years ahead of PETT, the clinical world swiftly adopted the SPECT scanner. I suspect that the swift adoption was due to inventors of the SPECT scanner being practicing doctors who understood the nitty-gritty of the clinical world. (Marcus Raichle (personal communication) comments that, because inventors of PETT were primarily basic research scientists, they commonly heard feedback from clinical practitioners that the prototypes were hard to use.) In 1963, Kuhl and his colleagues at the University of Pennsylvania developed their first proof-of-concept SPECT scanner (Kuhl and Edwards 1963). In 1968, they completed the Mark III scanner, the first SPECT scanner that was designed to be integrated into surgery planning procedures and ready to be commercialized (Kuhl and Edwards 1970). Kuhl himself was a practicing doctor. In 1973, a team at University of Aberdeen reported their own SEPCT scanner – the Aberdeen Section Scanner, based on Kuhl’s ideas (Bowley et al. 1973). Unlike the Mark III scanner, which was a dedicated brain scanner for speed and cost reasons, the Aberdeen Section Scanner was a general-purpose, whole-body scanner. Like the University of Pennsylvania
team, the Aberdeen team included practicing doctors, integrated the scanner into their clinical practice from the start, and their new scanner was ready to be commercialized.

5.2.3 PETT

In order to attract the interest of other researchers, academic and clinical institutions, and commercial manufacturers, researchers at WUSTL needed to persuade them that their new technique was not only better than the existing competitors, but also 4-million-dollars better than them. They kept failing. It is perhaps unsurprising that they faced the heaviest pushback from researchers already aligned with SPECT, given the similarity between PETT and SPECT.

The following is a reconstruction of Budinger’s and Keyes’s research teams’ critique (Budinger et al. 1977; Keyes Jr et al. 1977):

<table>
<thead>
<tr>
<th></th>
<th>X-ray CT</th>
<th>SPECT</th>
<th>PET</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution</td>
<td>Very High</td>
<td>Low</td>
<td>Low, 30% better than SPECT</td>
</tr>
<tr>
<td>Attenuation Correction</td>
<td>Good</td>
<td>OK (Correctable)</td>
<td>Good</td>
</tr>
<tr>
<td>Uniformity</td>
<td>N/A</td>
<td>Bad (Correctable)</td>
<td>Good</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>Good</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>General Purpose</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
</tr>
<tr>
<td>Convenience</td>
<td>Good</td>
<td>Medium</td>
<td>Low</td>
</tr>
<tr>
<td>Cost</td>
<td>Cheap</td>
<td>Affordable</td>
<td>Prohibitively High</td>
</tr>
</tbody>
</table>

Table 5.1: Comparison Table of Imaging Techniques

Here, as reconstructed in Table 5.1, Budinger and his colleagues, researchers at Berkeley already involved in SPECT research, acknowledged the theoretical advantages of PETT. They accepted that, for example, PETT produced at least a 30% improvement in spatial resolution (6 mm versus
They accepted that PETT could deal with tissue scattering and absorbing (attenuating, in general) gamma rays easily, while SPECT needed algorithms to correct attenuation. PETT had uniform depth response, while SPECT did not: SPECT suffered from significant signal loss if the signal came from deep inside the body. (The uniformity criterion does not apply to X-ray because X-ray does not come out of the body but passes through the body.) The list went on. Budinger and his colleagues, however, pointed out that those theoretical advantages failed to turn into any practical use. SPECT was already good enough in answering existing questions and there was room for SPECT to improve (Budinger et al. 1977). There did not seem to be a need to invest four million dollars into a new and unproven instrument. Keyes and his colleagues, another team at the University of Michigan, made similar claims. Keyes and his colleagues acknowledged, again, SPECT suffered from non-uniformity of depth response. Although attenuation and depth correction for SPECT was available and mostly sufficient for practical purposes, they acknowledged that PETT won out by the second and the third criteria. However, partially due to the commercial maturity of SPECT, they pointed out that SPECT won out in all practical matters. It was easier to use and much cheaper. They gently asked the same question to the PETT researchers: give us a reason to use this new, unproven instrument (Keyes Jr et al. 1977).

The response from WUSTL was underwhelming. Michael E. Phelps, who came up with the mechanical design and the algorithm for PETT, responded with an article named "What is the Purpose of Emission Computed Tomography in Nuclear Medicine?", although he never fully explained the purpose. In the article, he repeated PETT’s theoretical improvements over competitors that Budinger and his colleagues and Keyes and his colleagues all agreed with. Phelps then argued that PETT did have new practical uses. Phelps pointed out that PETT could generate tomographic images of the concentration of radioactive isotopes in an organ with high resolution and accuracy (Phelps 1977, p. 400), although he did not fully specify how PETT could answer questions in ways that existing, mature medical imaging techniques could not. Phelps acknowledged that SPECT had the advantage of working with longer half-life, commercially available radioactive isotopes while PETT
required an expensive cyclotron to manufacture radioactive isotopes on site. In order to deal with
the prohibitive cost, Phelps suggested that researchers could save the cost of cyclotron by adopting
PETT to using other, commercially mature, longer half-life radioactive isotopes that did not need
to be manufactured on site (Phelps 1977), although, again, he failed to explain what kind of new
answers PET could give with those long half-life radioactive isotopes. Lastly, he gestured that,
someday, PETT might be able to capture the dynamics of biological processes by taking a series
of rapid scans, although he did not have a concrete proposal for either such a scanner design or a
research agenda (p. 41).

Phelps’ response was not a success. If we look at the number of publications involving PETT, by
1977, it stayed at a very low level. Three articles were published in 1977 using the latest PETT

In short, surrounding PET, researchers found the following pros and cons:

<table>
<thead>
<tr>
<th>Pro</th>
<th>Con</th>
</tr>
</thead>
<tbody>
<tr>
<td>Better Attenuation</td>
<td>Prohibitive Cost</td>
</tr>
<tr>
<td>Better Resolution</td>
<td>No New Questions Answered</td>
</tr>
<tr>
<td>Depth Uniformity</td>
<td>Future Advantage Unclear</td>
</tr>
</tbody>
</table>

Table 5.2: PET’s Pro & Con

The summary table 5.2 should give us a general overview of the advantages and disadvantages
of PETT as they stood in 1977. First, PETT had advantages in central engineering norms. In the
medical imaging industry, attenuation, resolution, and depth uniformity are central norms at the
time and still now. It matters to have signals that can penetrate tissue, are relatively uniform despite
the distance traveled, and allow for discrimination between objects or fine details within an object.
Second, PETT had a scientific disadvantage. Although PETT did not answer questions themselves – researchers did, the instrument was unable to produce data that could facilitate exciting, new answers. Images with marginal increase in accuracy and resolution did not help researchers all that much.

Third, PETT had a serious pragmatic disadvantage. It took millions of dollars to build and maintain one PETT scanner and its associated cyclotron. Incorporating a new instrument into the existing academic or clinical routine would not be cheap either. For example, redesigning the diagnosis procedure, retraining the doctors, etc. were financially expensive and slow. PETT could be costly in another way. PETT could mean high opportunity cost to researchers and clinical practitioners. If researchers could improve existing instruments down the road (as Budinger and his colleagues and Keyes and his colleagues believed they could), then researchers could achieve the same performance with SPECT at a much lower cost, without investing in an expensive, new instrument with an uncertain future. Here, researchers were factoring future development, adoption, and opportunity costs into the current calculation. All in all, many believed that the costs outweighed the benefits. Raichle, a research team member at WUSTL, looked aback and remarked that the research team was, for a while, at a loss:

The idea that it would simply replace nuclear medicine brain scans was short-lived. There remained much work to be done to understand how access to a whole new range of in vivo measurements would inform us about basic biology as well as ultimately aid us in managing and treating patients (O’Connor 2013, p. 117).

To salvage their research and their new instrument, the team at WashU walked on a path similar to the team at MGH. First, they tried a variety of new directions while trying to improve the detection instrument as a good instrument in general, following standard norms. Second, they somewhat accidentally identified a neglected norm and started to reshape the competitive landscape: if PETT was not competitive in answering old questions and on the existing set of norms, maybe PETT could
become competitive on a neglected set of norms and answer new questions. And the turnaround story began. The next iteration of PETT developed at WUSTL improved upon its shortcoming of a really long scan time. The PETT scanner is a “tomographic” scanner and hence the “T” in the acronym. It is called a tomographic scanner because it is able to produce three-dimensional images of a scanned object. But in 1974, the prototype PETT III scanner was only able to scan one two-dimensional slice of a three-dimensional object at a time. To produce the full three-dimensional image, the PETT III scanner had to scan one slice, move the detectors, scan the next slice, and keep repeating the process. The scanning process was long. With the PETT III scanner, researchers needed 20 separate two-dimensional scans for the full three-dimensional picture, with each scan taking 2 minutes (Hoffman et al. 1976, pp. 495-496). Here, I am talking only about the scanning time, not including the time to mechanically move the detectors, to post-process the data, etc.

To shorten the unusually long scanning time, researchers at WUSTL developed the PETT IV scanner in 1978, which could scan seven slices concurrently at a time. The PETT IV scanner had one major advantage and one major disadvantage. On the one hand, as expected, researchers were able to reduce the scan time by more than half. A full scan by the PETT IV scanner took roughly 20 minutes, while a partial scan of seven slices took roughly 6 minutes (Ter-Pogossian, Mullani, Hood, Higginls, et al. 1978, p. 483). In comparison to more than 40 minutes by the PETT III scanner, the PETT IV scanner was a significant improvement. However, the PETT IV scanner was also a significant setback in terms of spatial resolution. It’s Full Width at Half Maximum (FWHM), a measurement of spatial resolution, was at best 12mm and normally 15mm, while the PETT III scanner’s FWHM was 7.5mm at the center (Budinger et al. 1977, p. 23). While I will explain the concept in much depth in the next section, FWHM can be intuitively understood as how close two objects or signals are before the detection instrument cannot tell the two objects apart. Increasing FWHM from 7.5mm to 12mm means that PETT IV needed signals to be further apart to distinguish them, and hence reduced the level of detail that it could produce. Cox et al. pointed out that, one of the fundamental trade-offs in medical imaging was the trade-off between spatial and temporal resolution. The total
amount of data to be detected and processed is the amount of data per each scan times the number of scans in a given time interval. To oversimplify things a little bit, higher temporal resolution means making fewer scans within a given time interval. Because the cameras can only detect a certain number of gamma rays and the back-end computer can only process a certain number of data, to take many scans in a short time requires one to reduce the amount of data per scan (Cox, Hill, and Mullani 1974, p. 34), which leads to poor spatial resolution of each scan.

I must point out that researchers developed the PETT IV scanner not primarily to capture fast changing biological processes. Like Kuhl and his colleagues’ MARK III SPECT scanner (Kuhl, Edwards, and Ricci 1974), researchers at WUSTL improved upon temporal resolution so as not to inconvenience themselves or the patients. Previously, with PETT III, a full scan took a full hour from the beginning to end, which was an inconvenience to both researchers and patients. In comparison, a two-dimensional X-ray took seconds. Researchers felt the pragmatic pressure to make faster scans.

With the PETT IV scanner in hand, Ter-Pogossian took the development the PET scanner in a less explored direction. In the 1970s, the general trend of medical imaging was to develop general-purpose, high spatial resolution imaging devices that could scan any part of the body and detect tumors or abscesses in any organ (Mallinckrodt Institute of Radiology 1975; Keyes Jr et al. 1977). The trend, in other words, involved two goals: first, a pragmatic design goal to ensure the imaging device could be as versatile as possible; second, a normative prioritization to emphasize spatial resolution and, if there are trade-offs, try to preserve and improve on spatial resolution. The research team at WUSTL had a different idea. Raichle, a junior member of the team at the time, suggested that the team should design the PETT scanner for an entirely different function. Raichle had some vested interest. While he had an M.D., his academic interests were not in tumor diagnosis or clinical practice. He primarily worked on cerebral metabolism and was acquainted with using lesion studies to figure out the neurological basis of psychological functions (Raichle 2014). He had two reasons to
design the PETT scanner to image the brain (Mallinckrodt Institute of Radiology 1977): first, a large number of brain diseases (e.g. stroke, etc.) affect the brain by affecting cerebral metabolism; second, there had been no good tools to image the normal function of the human brain in vivo. Raichle in other words pushed for a fundamental repurposing of the PETT research program. Instead of trying to build and find uses for a diagnostic imaging device, Raichle was interested in building an imaging device to study the normal function of the brain. Ter-Pogossian, the leader of the research team, long had interest in the brain himself, primarily in stroke. With the input from Raichle, Ter-Pogossian noticed the potential value of a device focused on the brain and designed to further improve temporal resolution to capture the fast-changing cerebral metabolic processes (Freeman 1988).

In 1979, just a year after the PETT IV scanner, the research team at WUSTL came up with a radical new design – the PETT V scanner. It was a radical new design because, with the PETT V scanner, researchers at WUSTL abandoned the ambition of previous iterations of PETT scanners. Researchers previously intended the PETT scanners as general purpose scanners. They designed and used PETT III and PETT IV scanners to scan the human brain, heart, lung and animals. Keys et al. pointed out that flexibility to scan any part of the body was a desirable virtue of an imaging technique (Keyes Jr et al. 1977). A versatile imaging device could greatly save cost. After all, buying separate imaging devices for different purposes could soon accumulate the initial purchase and ongoing maintenance bills. Now, researchers at WUSTL designed the PETT V scanner as a special purpose scanner devoted to the human brain. Abandoning the flexibility resulted in significant improvement in temporal resolution. With only the head to cover, the new PETT V scanner was significantly smaller and the scanning time was shortened. The theoretical minimal scan time for seven slices took merely one second (Ter-Pogossian, Mullani, Hood, Higgins, et al. 1978, p. 540). At the same time, researchers managed to improve spatial resolution. Because mathematically spatial resolution and the amount of noise per each scan partially depend on the size of the scanner, shrinking the size of the scanner also greatly improved the spatial resolution. The theoretical spatial resolution of the PETT V scanner was back to 7 mm, slightly better than the PETT III scanner (p. 540).
Two years later, the research team finished the PETT VI scanner. Researchers adopted a new detector design based on a new crystal and further improved computational circuit design in the PETT VI scanner. In comparison to SPECT scanners, the PETT VI scanner was more sensitive and could scan much faster. Researchers changed the crystal to detect gamma rays from NaI(Ti) to 288-Cesium Fluoride. 288-Cesium Fluoride’s decay time was only 2% of the original NaI(Ti)’s decay time, speeding up the crystal’s response time to the arrival of gamma rays. Coupled with better computational circuit design, researchers were able to further crank up the temporal resolution. In the end, the resolving time, the amount of time needed for crystals to respond, amplifiers to amplify the signal, and computers to calculate all combined, was 3.4 times shorter than other alternatives (Ter-Pogossian et al. 1982, p. 129). With those improvements, researchers pushed the theoretical full scan time down to only one second (Yamamoto, Ficke, and Ter-Pogossian 1982, p. 529), although the more realistic scan time was roughly 20-40 seconds (Ficke et al. 1982, p. 478). The PETT VI scanner was the fastest brain imaging device at the time (Mallinckrodt Institute of Radiology 1980).

In 1981, four years after Phelps’ earlier suggestions, Ter-Pogossian reversed Phelps position and published his insights. In this article named “Special Characteristics and Potential for Dynamic Function Studies with PET”, unlike Phelps, Ter-Pogossian insisted on using the short half-life radioactive isotopes at the high cost of installing and operating a cyclotron. Most commercially available radioactive isotopes ready for PET and SPECT scanners at the time were perfusion tracers. Perfusion tracers do not participate in biological processes. They just diffuse into tissue through the blood flow. Researchers can use them to differentiate soft tissue from bones, air, or to differentiate normal tissue from abnormal tumors, etc. Those perfusion tracers help researchers highlight anatomical structures. Those short half-life radioactive isotopes are radically different. Because Carbon, Nitrogen, Oxygen and Fluorine are primary participants of biological processes, by tracking their location and level of concentration, researchers could track biological processes. For this reason, Ter-Pogossian called them the “physiologic” radioactive isotopes. Thus he remarked:
The essential value of PET, now a full-fledged technique, consists in the fact that this methodology permits in a number of instances the in vivo, regional, and noninvasive investigation of many biochemical processes essential to life (Michel M. Ter-Pogossian 1981, p. 13).

In other words, Ter-Pogossian forcefully pointed out that PETT was not special because it was incrementally better than other techniques in a crowded, competitive landscape. Instead, PETT was good partially because it could do something totally new. Ter-Pogossian then pointed out that tracking biological processes had a serious obstacle. A large number of biological processes essential to life and cognition change rapidly. Researchers could not tell anything if a full scan took 40 minutes. In order to study those biological processes, he needed new scanners and new methods that were fast (p. 14). The PETT VI scanner was the only fast medical imaging devices on the market and the only one to answer those new questions. And coupled with the metabolic ("physiologic") tracers, PETT was the only technique to study the physiological basis of human cognition.

Precisely a decade after the papers announcing their invention of the PETT scanner, the research team at WUSTL thought they had a definitive answer to why fellow researchers should use PETT in academic research. Ter-Pogossian, in 1985, reiterated the reasons that he had already listed in his 1981 paper and compared PETT with other medical imaging techniques on the market (Michel M. Ter-Pogossian 1985). First, with the help of Carbon-11, Nitrogen-13, Oxygen-15, and Fluorine-18, researchers with PET could track metabolic events inside the human brain. In other words, unlike CT and SPECT at the time, which could only provide anatomical information, PETT could provide functional information – where and how metabolic events took place. Second, PET could provide dynamic information. In other words, unlike CT and SPECT at the time, which could only produce static images, PETT could track the rapidly changing metabolic events in real time.

Here, Ter-Pogossian highlighted a new competitive ground: many questions about the living, normal biological processes, especially those about the brain, were not answered and were previously
limited by the detection instruments available to researchers. PETT, after a few iterations, could be a new, forceful tool to answer those questions, because it was fast enough to study to fast-changing biological processes. At the same time, researching increasingly faster detection instruments was important. Increasing the speed of imaging had always been a norm in detection instrument design. But speed was then directly tied to studying critical normal functioning of the brain, the norm also received more attention and was given priority.

The academic acceptance of the PETT scanner and the PET technology is visible through the tracing of journal publications in the 20 years after 1977. Here, to produce Figure 5.1, I use the same dataset generated in the previous chapter.

The dataset captures the journal articles published using the PET technology to study the brain. I further added the disciplines of the journals that those articles appeared in. In particular, the gray bar is the number of publications in neuroscience. The orange bar is the number of publications in radiology. The bright blue bar is the number of publications in endocrinology and metabolism. The trend of article publication tells at least two stories. First, by around 1980 and 1981, the number of articles using PET exploded. This is evidence for general interest from the academia. Second, while the early papers in 1977 and 1978 were exclusively “technology” papers on developing PET equipment and technologies, by around 1984, articles primarily focus on using the technologies to solve scientific problems, especially on neuroscience.

While I am presenting correlative data here, I think there is evidence to think that there is a causal relationship between WUSTL team’s efforts and the adoption of PET in academic, especially neuroscientific studies. For decades, PET was primarily an engineering development: it was a new detection instrument based on a specific radiological principle. Researchers, including those at MGH and at WUSTL, were trying to find academic and scientific uses for the device. The breakthrough came from the WUSTL team to turn the device to study normal phenomena of the brain and focus on the norm of increasing temporal resolution, utilizing its new and fast device,
Figure 5.1: Distribution of Journal Categories Publishing PET Research
PETT VI. The dataset suggests that, around the same time, researchers not only started to use PET, but also to use it for the intended, neuroscientific inquiries, instead of merely engineering development. Given the unique development at WUSTL, I suggest that it is plausible to conclude that it is the WUSTL research team who convinced their colleagues to use PET and solidified PET’s role in neuroscience. They convinced their colleagues because PET opened up a new field with both exciting opportunities and no existing competitors. PET enables researchers to study the normal cognitive functions. In studying human normal cognitive functions, although accuracy and resolution do matter, a new virtue, increasing speed, matters even more. Without speed, there is no way to image the fast-changing processes in the brain. This is precisely why the research team at WUSTL used 288-Cesium Fluoride as the new crystal in the PETT IV scanner. In comparison to NaI(Tl), 288-Cesium Fluoride produced slightly noisier pictures, making it slightly more difficult to tell metabolic changes in the image. But it was fast – more than twenty folds faster. Speed was the priority when we were talking about biological processes in the brain. They convinced their colleagues primarily because a new engineering development could solve new and more scientific questions.

5.3 Understanding Resolution

The WUSTL team, during their pivotal years to create a viable technical and practical market for their nascent PETT, struggled with resolution. They wanted to offer more details in their scans (alluding to the desire to increase spatial resolution) and they wanted to offer faster scans (alluding to the desire to increase temporal resolution). But as Cox et al. pointed out, other things equal, there normally is a trade-off between spatial resolution and temporal resolution. The WUSTL eventually emphasized increasing temporal resolution and found widespread adoption of their new detection instrument.

In this section, I use the historical episode as source material for philosophical analysis. The interest in resolution has a long history in philosophy. In particular, I call attention to Hacking’s discussion
of resolution and I rely on the WUSTL team’s history to pick up from Hacking’s philosophical discussion. I ultimately argue that the need for more resolution in science and engineering is not just merely a need for more information per se. It is about the desire to have detection instruments that can offer different levels of information aggregation. To use an example used by Hacking, the telescope is such an instrument. If we look at the moon through our naked eyes, we see an area of brightness in the sky and, if we have good eyesight, we may even see uneven shades of brightness and darkness. However, if we use the telescope, we see a different picture: we will see mountains and valleys on the moon. If the former, naked-eye view inaccurate? No, it is totally accurate as an aggregate of sunlight reflected by the moon’s surface. But for various reasons, including curiosity about the universe, we want to have information at a different level of aggregation. The telescope allows us to separate and de-aggregate information and see the mountains and valleys on the moon.

Resolution is a measurement of the level of aggregation. In particular, for spatial resolution, it measures how spatially close two pieces of information could be for the detection instrument to still acquire both and distinguish them apart, instead of aggregating them. For temporal resolution, it measures how temporally close two pieces of information could be for the detection instrument to still acquire both and distinguish them.

In this section, I first review the philosophical literature, especially Hacking’s discussion of resolution, before I dive into the philosophical significance of spatial resolution and temporal resolution separately.

5.3.1 Continuing Hacking’s Unfinished Project

The role of resolution in scientific instrumentation receives significant attention from Hacking. Although he makes important inroads in drawing our attention to the exceptional importance of resolution to a proper epistemology, he does not explore the issue in depth. Hacking discusses the issue of resolution when he recounts the early history of the microscope. One challenge was aberration, namely, that imperfectly crafted lens would introduce interfering lights and cause the
image to look like one through a kaleidoscope (Hacking 1983, p. 193). The challenge of aberration is partially a challenge of low signal-to-noise ratio, as discussed in Chapter 3.

Another challenge with the early microscope, relevant to this chapter, was magnification. Hacking points out a significant difference between magnification and resolution. Despite its length, I am quoting the entire paragraph due to its significance (p. 195):

Abbe was interested in resolution. Magnification is worthless if it ‘magnifies’ two distinct dots into one big blur. One needs to resolve the dots into two distinct images. It is a matter of diffraction. The most familiar example of diffraction is the fact that shadows of objects with sharp boundaries are fuzzy. This is a consequence of the wave character of light. When light travels between two narrow slits, some of the beams may go straight through, but some of it will bend off at an angle to the main beam, and some more will bend off at a larger angle: these are the first-order, second-order, etc., diffracted rays. Abbe took as his problem how to resolve (i.e., visibly distinguish) parallel lines on a diatom (the tiny oceanic creatures that whales eat by the billions). These lines are very close together and of almost uniform separation and width. He was soon able to take advantage of even more regular artificial diffraction gratings. His analysis is an interesting example of the way in which pure science is applied, for he worked out the theory for the pure case of looking at a diatom or diffraction grating, and inferred that this represents the infinite complexity of the physics of seeing a heterogeneous object with a microscope.

In this paragraph, Hacking recounts the struggle in the early designs of the microscope. The microscope works by passing light through objects, often microorganisms. Each surface spot of objects will absorb some of the light and reflect the remaining light. The reflected light is useful because it carries information about the surface of objects. But the reflected light waves will interfere with one another, making it difficult to tell where each light wave comes from. With a microscope
that cannot separate light waves based on its origins, we see blurriness. With a microscope that can, we see minute details of the object.

In the passage and in the context, Hacking distinguishes the difference between mere magnification and increased resolution: it is possible to enlarge a picture but without more details. This is the case of mere magnification. It does not provide more information about the microscopic world of interest to researchers. But designers of the microscope want magnification (enlarging the picture) and increased resolution (increased minute details). Hacking’s discussion of resolution ends at highlighting its distinctiveness and importance.

Researchers following up on Hacking’s discussion of resolution is limited. Resolution, in a few ways, is distinct from other common measurements that philosophers normally talk about. Accuracy, as I mentioned in the introduction, is a complicated notion. Tal identified five different notions of accuracy (Tal 2015, p. 1084). But in general, accuracy is a measurement about how our measurement of a value deviates from its true value, or how big the measurement error is. Precision, likewise, has many definitions. But commonly, precision is about the agreement between measurements (Teller 2013). Resolution, as recounted by Hacking, is neither. Resolution, especially spatial resolution, is perhaps most similar to the discussion of perceptual clarity. One theory of perceptual processing of distant objects involves not only point estimates of where the distant objects are and their features, but also an uncertainty distribution. Clearer objects to the eyes have a smaller uncertainty distribution (Vance 2020). As it will become clear, resolution is assessed along the same line as clarity. But clarity is the phenomenal experience of seeing objects. Resolution is a measurement of detection instruments.

The short review suggests that the notion of resolution, as a measurement of features of detection instruments, requires more philosophical investigation and clarification. In this chapter, I use the WUSTL team’s experience to ground my philosophical examination of resolution in depth and to offer a philosophical account of resolution. I intend to demonstrate the scientific and engineering
significance of resolution, provide philosophical and scientific definitions of both spatial and temporal resolutions and their associated engineering norms. In the next section, I borrow from evolutionary metaphors to identify one mechanism that researchers relied on to improve on resolution and find market opportunities for their nascent detection instrument. While the history is focused on merely one detection system – Positron Emission Tomography and its predecessors, I intend the philosophical accounts of both the norms and the mechanisms to apply to other detection systems as well.

5.3.2 Resolution

Like signal-to-noise ratio and sensitivity, resolution is another key measurement of detection systems and increasing resolution is a common norm in detection system design. Resolution, in the practice of detection systems, is a measurement of whether and the extent to which a detection system can separate and identify two pieces of information as two pieces of information. One might wonder why or how a detection system can fail to separate two pieces of information. Such failures are extremely common in our daily phenomenal experience and in research, especially in cases of visual clarity and visual blurriness as mentioned by philosophers. For example, myopia patients cannot distinguish distant information. To myopia patients, any details beyond a certain range look blurry, and they cannot tell, for example, that two far-out people are actually two people. The struggle to distinguish distinct information is not just common to visual experience. Touch is another useful example. Different parts of human hands have different receptive field sizes (Gardner 2010). Our fingertips are extremely sensitive to information about location. We can have sensory distinction of touching two points separated by, in some cases, merely 2 mm apart. But our palms can only sense two points separated by around 10 mm apart.

Because resolution is an umbrella measurement of how a real difference appears to us as different, researchers have identified different kinds of resolution that they care about. After all, there are many kinds of real differences. The following is a non-exhaustive list:
• Frequency Resolution: researchers want to measure how a detection system can separate
different frequencies from each other (despite that overlapping signals may interfere with
each other)

• Energy Resolution: researchers want to measure how a detection system can separate different
energy levels from each other

• Spatial Resolution: researchers want to measure how a detection system can separate infor-
mation from distinct spatial locations from each other

• Temporal Resolution: researchers want to measure how a detection system can separate
information from different temporal points from each other

Because of the available kinds of resolution in play, and because each kind is often measured and
improved in distinct ways, in this chapter, I will focus on Spatial Resolution (colloquially known as
“Resolution”) and Temporal Resolution (colloquially known as “Speed”).

5.3.2.1 Spatial Resolution

The main challenge with spatial resolution is that no detection system (including the human eye)
can detect natural objects exactly as they are, without any distortion. Any detection system will
spatially distort natural objects somewhat. In order to increase spatial resolution, researchers often
need to identify patterns of distortion so they can a) post-process the information based on known
distortion patterns and b) better find ways to reduce distortion.

In order to better identify the patterns of distortion and limit the distortion, researchers come up
with mathematical means to describe the distortion. They use a wide variety of mathematical
characterizations, such as Point Spread Function (Prince and Links 2014, p. 25), Line Spread
Function (p. 61), etc. (The details of Point Spread Function were discussed in Chapter 3.) Most of
the common characterizations follow the same approach: researchers use their detection systems to
detect a known source and then use a mathematical function to describe the pattern of distortion.
With the distortion function available, researchers can evaluate a detection system’s spatial resolution through a measurement known as Full-Width at Half-Maximum (FWHM) (Prince and Links 2014, p. 63). Here, in Figure 5.2, we illustrate two curves – distortions created by detecting two spatially

Figure 5.2: Full-Width at Half-Maximum
separate sources. If the two objects are sufficiently far apart (as in subfigure (a)), then the two curves are identified as separate. If the two objects are too close (as in subfigure (d)), then the two curves overlap and we may identify them as one coming from a single information source. But what defines “too close” or “too far”? How should researchers draw a line between a continuum between subfigure (a) and subfigure (d)? Like perceptual clarity, the line is partially drawn through the distribution of uncertainty (Vance 2020). If the distribution of uncertainty of two objects overlaps significantly, then it is determined that the detection instruments could not be separated from each other. The notion of “significance” is vague and often operationalized as FWHM. FWHM, or Full-Width at Half-Maximum, is the measurement of the width of the curve at half of its height, or the FWHM line at each curve’s waist in Figure 5.2. If the separation between two curves is under FWHM (such as subfigure (c) and (d)), then we cannot tell if they are actually separate curves and hence coming from just a single information source. If the separation between two curves is more than FWHM (such as subfigure (a) and (b)), then we treat them as actually separate.

Figure 5.2 also provides a concrete example of why resolution is also a measurement of the level of aggregation. In subfigure (c) and subfigure (d), when the detection instrument cannot separate the two signals, they are detected as a single peak. In other words, when the detection instrument cannot distinguish the two, the two are aggregated. The term “aggregation” is unfortunately vague here because detection instruments often aggregate different signals in different ways, sometimes merely additively, sometimes taking average, but at other times, in more complex ways. The philosophical point here does not rely on a precise definition of aggregation: if a detection instrument cannot distinguish objects 2 mm apart, it aggregates information at the 2mm level or below. Wanting to increase resolution, say, from distinguishing objects 2 mm apart to distinguishing objects 1 mm apart, is, in other words, a desire to aggregate information at merely 1 mm level or below, instead of 2 mm level or below.
In the historical section, I have reviewed the struggle with FWHM faced by the WUSTL team. From the transition to PETT III to PETT IV, the WUSTL team slashed the image acquisition time by more than a half, but as a trade off, the resolution worsened. From PETT III to PETT IV, the FWHM increased from 7.5 mm to 12-15 mm. It means that PETT III could identify two spots 7.5 mm apart as two separate spots, but PETT IV may only record one spot. Fortunately, at the grand scale of history, in general, the FWHM of PET scanners was decreasing. I hope Figure 5.3 can further help to demonstrate the decreasing FWHM, and hence increasing spatial resolution, had been a consistent desire and trend in PET research.

“Resolution,” similar to sensitivity, is impacted by a number of effects. Unfortunately, unlike sensitivity, there is no good list of major factors that contribute to a detection system’s resolution or a comprehensive mathematical formulation. Fortunately, in the history covered in this chapter, we have already witnessed a number of considerations that influence resolution. Let us focus on the
example of minimal response area. Minimal response area is one example of applying FWHM in real life. It says that, any source of information coming from a given region cannot be separated spatially from another source from the same region. It is because the FWHM is bigger than the minimal response area. For example, in the early years, as retold in Chapter 4, Lawrence and his colleagues at Berkeley had only rudimentary Geiger-Müller detector and could only report radioactivity levels at the scale of organs: “liver,” “bone” etc. In other words, any radioactivity coming from the liver cannot be spatially separated from other radioactivity also from the liver. Researchers at MGH were more intentional in reducing the minimal response area. They intended to find fine-detailed ways to detect tumorous tissue in the brain. They hence explicitly desired a new Geiger-Müller detector with a tip.

5.3.2.2 Temporal Resolution

Temporal Resolution, in definition and in measurement, is extremely similar to spatial resolution. Spatial resolution is defined as the ability to distinguish two spatially separate objects or events as separate. Similarly, temporal resolution is defined as the ability to distinguish two temporally separate objects or events as separate (Prince and Links 2014, p. 69). Similarly, temporal resolution can be measured by constructing a Point Spread Function and by FWHM. In Figure 5.4, the graphic “distance” between x1 and x2 is not labeled. In the spatial resolution subsection, we assumed the distance was spatial distance. In this subsection, we can re-use the same graph because the distance could equally refer to temporal distance.
5.4 Niche Construction

The history of PET at WUSTL saw a happy ending: the research team found that many new questions on the normal functioning of the brain and the body can be answered by a high-temporal resolution.
PETT and only by their new PETT. And the academic market reacted with enthusiasm to integrate the new device into their academic research. What can a philosopher learn from the history? First of all, increasing temporal resolution, as mentioned in the previous chapter, is a desirable and important but at the time, somewhat neglected norm. Second, the team at WUSTL seems to be very effective in exploiting untapped scientific and technological opportunities. Is it possible that their “opportunistic” prioritization of the untapped norm of increasing temporal resolution is generally applicable in the understanding of patterns of scientific instrumentation and tool development?

I think the answer is a positive, because their exploitation of opportunities fits very well with a generally accepted evolutionary mechanism: niche construction. As I have mentioned in Chapter 2, scientific instrumentation has many features similar to the evolutionary process. Tool development is a competitive process: researchers at different institutions often come up with overlapping and competing tools. Just like competition in the natural world, competition in tool development often leads to rewriting the rules of the game – modifying the regulating norms or finding new norms along which research programs and technologies should be assessed, and shaping the trajectory of tool development. In the case of Positron Emission Tomography, competition led researchers at WUSTL to find an unoccupied norm, increasing speed, and elevated the role of speed in neuroscience.

Niche construction theory is an evolutionary theory that emphasizes the role organisms play in shaping the competitive landscape and describes the mechanisms of actively shaping the competitive landscape. In the 80s, following Lewontin (Lewontin 1982, p. 160) and Waddington (C. H. Waddington 2008, p. 262), Odling-Smee was one of the first biologists to propose a systematic account of niche construction. The niche, according to Odling-smee et al., is “the sum of all the natural selection pressure to which the population is exposed” (Odling-Smee, Laland, and Feldman 1996, p. 40). Niche construction, then, is when an organism modifies the relationship between parts of the organism and parts of the environment by actively changing parts of the environment, by perturbing the environment, or by relocating to a more favorable environment (p. 41). In other
words, niche construction theory challenges scientists’ previously preconceived description of the dynamics between organisms and the environment. Previously, scientists often emphasized how organisms are passively selected. Odling-smee and other niche construction theorists emphasize how organisms select, modify, and create the environment in which they compete against others.

Niche construction theory provides a number of important insights for us to better understand and explain the mechanism of scientific instrumentation and tool development. First, the effect of niche construction is persistent. Niche construction shapes the environment in the long-term and creates a secondary channel of inheritance - ecological inheritance (Odling-Smee, Laland, and Feldman 1996, p. 20). One straightforward example is egg-laying. When the female deposits an egg, the female normally deposits the egg on or near the food source. The offspring inherits not only its parents’ genes, but also the favorable environment. Similarly, nets of social bees, wasps, and other animals become the selection pressure for many nest regulatory behaviors (p. 62).

Second, niche construction, through shaping the competitive landscape, shapes other features of the niche constructing organisms and other species (Odling-Smee, Laland, and Feldman 1996, p. 30; Odling-Smee 1988, p. 99). The environment is heterogeneous and fluctuates all the time. As organisms niche-construct the environment, the selective pressure changes accordingly and what counts as adaptation will fluctuate as well (Odling-Smee 1988, p. 90). For example, birds probably first evolved feathers for thermoregulation. But once feathers were exaptated for flights and changed the competitive landscape, other body parts (such as chest muscle) went through secondary evolution to better facilitate flight (Gould and Vrba 1982, p. 11). For another example, beaver’s well-known dam-building behaviors change the water landscape and shape the behaviors of other animals who rely on the water landscape.

To reformulate niche construction theory’s insight, we have the following two points for tool development. First, tool development can seek the pragmatically beneficial norms to govern its trajectory and hence have a lasting impact on the competitive landscape. In previous chapters, I
have discussed norms such as increasing signal-to-noise ratio, increasing sensitivity, increasing spatial resolution, and increasing temporal resolution, etc. It is often impossible for researchers to improve their instruments along more than one norm. For example, in this chapter, I discussed the common conflict between increasing spatial resolution and temporal resolution. While the content of those norms is fairly fixed, the relative importance and priority of engineering norms is often a social and pragmatic issue. In the history of PET, researchers at WUSTL, once realized that they might not win out in the competition for better accuracy or better spatial resolution, found out that there was not sufficient competition for speed, especially speedy scans of the brain. Hence they focused on the developed for speedier instrument, developed a new research program and a new set of questions around speedy imaging of the brain, elevated the significance of speed in the research community, and also elevated their own research program.

Second, the shifting competitive landscape encourages a wholesale redesign both of the tool development at hand and other ones interested in competing in the environment. Just like birds that went through wholesale redesign to better facilitate flight, PET went through wholesale redesign to focus on speed. Researchers abandoned the general-purpose PETT IV scanner and came up with the special purpose PETT V scanner exactly for this reason. Replacing the slow NaI(Tl) crystal with 288-Cesium Fluoride crystal served the same purpose. The success of PET, which offered fast cerebral scans, led to later generations of imaging techniques that focused on speed too. Researchers also created an entirely new field, neuroimaging, with almost completely new goals, questions and methodologies. The goal of neuroimaging is to study the biological processes underlying cognition in the brain. The regulative norm in neuroimaging is precisely speed. Biological processes are complex and rapidly changing. Only fast imaging techniques could help us reveal the underlying mechanisms. Other new tools later to compete in the field had to offer similar or better speed too. Functional Magnetic Resonance Imaging (fMRI) which eventually replaced the PET scanner in studying the brain had to be significantly faster than PET to be competitive. The new focus on speed encouraged researchers to find novel applications of fast scanning in other fields too, such as fast

175
CT to monitor tumor motion or treatment delivery in real time (Thomas et al. 2014; Wu, Lian, and Shen 2012).

5.5 Conclusion

In this chapter, I briefly reviewed the efforts at Washington University in St. Louis in the 1970s to find popular adoption for PET scanner. Researchers at Washington University in St. Louis worked on and developed an instrument with inputs and inspirations from Massachusetts General Hospital, University of Pennsylvania, and other institutions. They very quickly found themselves at a critical conjuncture: the instrument was technically mature enough for adoption, but not academically or clinically mature enough – well coupled with academic questions or clinical procedures to compete against existing solutions. In the same market that researchers intended the instrument to occupy, more mature, more accepted solutions, such as X-Ray CT and SPECT, already existed. It was hard to convince other researchers, clinical practitioners, and institutions to purchase an unproven machine priced more than a million dollars, merely because the machine could produce marginally better images.

Competition forced researchers to pivot and find an unoccupied niche. In a somewhat accidental fashion, researchers at WUSTL found a breakthrough. Because of their interest in cerebral physiology and because of some accidental design choices, they came up with a brain-focused design that could perform a scan every few seconds instead of every half an hour. In the next few years, around 1980, researchers at WUSTL witnessed swift adoption of their scanner. The adoption was primarily due to meeting a latent and scientifically significant demand: seeing cerebral physiological changes in real time.

The development and pivoting process for researchers at WUSTL draws significant similarity with the niche construction process. Researchers at WUSTL do not compete against other competitors regarding the same set of norms (such as spatial resolution). Instead, somewhat incidentally, they
found an unoccupied niche – fast imaging of cerebral physiological processes, occupied the niche, transformed their instrument to better fit the niche, and saw great academic and commercial success. It is the changing the competitive landscape, finding a new norm (faster scanning speed), and eventually re-writing the competitive landscape that helped PET’s adoption.
Chapter 6

The Epilogue the 1970s to 1990s, The Role of Modularity

6.1 Introduction

In the previous chapter, I covered the major technical, scientific, social, and normative break-throughs of PET research at Washington University in St. Louis (WUSTL). Researchers at WUSTL accomplished not only a major technical breakthrough to image the brain fast enough to observe cerebral biological processes in real time. They also enabled and highlighted the scientific, social, and normative value of studying normal brain functions based on imaging cerebral physiological changes in real time. Based on my counting of academic publications using PET in the previous chapter, their research quickly gathered the interest of the broader research community.

But my narrative is incomplete. While I put together a narrative account of how researchers at WUSTL gathered interests in the academic community, I have not discussed how the availability of technology changed the academic community and how researchers used PET for novel research directions in the ensuing years. From the 1970s to the 1980s, and the 1990s, WUSTL researchers
witnessed the explosive growth of publications using PET from merely a handful every year in the 1970s to hundreds and thousands every year in the 1990s. But what kind of research programs and projects were the publications reporting on? The number of publications is a high-level and crude measurement of the academic community’s interest in and adoption of the new technology. It does not provide fine-grained information on how the academic community responded to a new technology.

At the same time, tracking the later years of PET is exponentially harder than writing about its earlier years. When researchers on PET published only a dozen papers every year, it is possible to go through every single article, track down how researchers discussed or argued with each other, and chronicle how researchers were using, improving, modifying, or criticizing the technology. But once the field explodes to hundreds of new publications every year in the 1980s, identifying key trends and directions in the field becomes extremely challenging. Trying to tell a single, narrative story focusing on one or a few teams would neglect the motivations and achievements of a wide range of other researchers pursuing their own products with a variety of motivations. Instead, we need to develop questions and techniques to answer them at a high-level of organization in the social structure of the field.

In this chapter, I write the epilogue using the bibliometric method. Instead of telling the narrative story based on a review of the full text of publications from a few research teams, I construct the story of how PET took off using the “meta-data” of publications – the dates, authors, affiliated institutions of publications, how one publication cites the body of literature, and other information outside the full text. By crunching the metadata, I advance two arguments. First of all, the field of PET has a “two-engine” structure. While scientific pursuits and interests drive the development of the field, the PET research community clusters around their choice specific technologies. While

4. The data analysis here is impossible without the support of the Digital Humanities Project at Washington University in St. Louis and other members at the Philosophy-Neuroscience-Psychology Program. In particular, I am extremely thankful for the help and contribution from Ronny Bass, Alex DiChristofano, Douglas Knox, David Gruskin, Luis Goicouria, and Steve Pentecost.
some subcommunities in PET research are formed based on scientific topics (such as neuroanatomy), many are formed around researchers’ choice of specific features of PET or their choice of radioactive isotopes to work with PET, such as using PET with carbon isotopes, with oxygen isotopes, and with fluoride isotopes. Second, I argue that the community structure is partially the product of PET’s survival through modularity. In order to survive and expand into exciting new research directions, researchers rely on modular modification of their instruments and create a sub-community to pursue new directions.

To make the argument for the role of modularity in forming the community structure, I divide the chapter into three sections. I begin with my method of putting the requisite bibliometric data together to study the body of literature on PET from the 1960s to the 1990s. I then discuss the results and their significance. I end with a theoretical discussion of modularity and community ecology, and I emphasize their importance in understanding scientific instrumentation.

### 6.2 The Bibliometric Method

To study more than 20 years of the development of a technology is a tremendous task. In order to process and extract information from large bodies of literature, researchers have proposed many computer-assisted, algorithm-based methods, such as automatically analyzing the textual content of bodies of literature and generating important themes and topics in bodies of literature or visualizing bodies of literature over their respective geospatial or temporal scales (Hughes, Constantopoulos, and Dallas 2015). In this chapter and dissertation, I mostly rely on the bibliometric research method promoted by de Solla Price (Price 1965) and systematically proposed by Garfield et al. in the 1960s (Garfield, Sher, and Torpie 1964). Their bibliometric research method is based on the core insight that citation is a chronological and intellectual link between bodies of literature (p. 1). If a paper A cites paper B, then it is reasonable to assume that researchers of paper A learned something from paper B. The relationship does not have to be between individual papers and perhaps is more fitting for bodies of literature: it is hard to tell what researchers of paper A learned from paper B.
from a single citation, without going through the detailed discussion in paper A. However, if paper A cites, for example, a substantial number of papers all related to evolutionary psychology, it is reasonable to infer that researchers of paper A engaged with the evolutionary psychology literature substantially.

Among many algorithms useful in the bibliometric method, I pay most attention to Main Path Analysis. Main Path Analysis is a structural approach to a temporally ordered network of citations proposed by Hummon and Dereian (Hummon and Dereian 1989) with a more efficient algorithm developed by Batagelj (Batagelj 2003). The goal of the algorithm is to identify “agenda setting” publications at each point of time in a chronically ordered set of publications. Here, “agenda setting” is arithmetically understood as publications that synthesize prior research and are then cited as authoritative by later research (Lucio-Arias and Leydesdorff 2008, p. 1951). Those papers are what the whole field “flows” through, such as Newton’s or Einstein’s papers. Although not every paper in physics engages with Einstein’s papers, they synthesize from earlier papers and become monuments that later papers have to deal with. In a way, Main Path Analysis is an attempt to identify key influencers like Newton’s or Einstein’s papers in any body of literature. Although it will miss out a lot of other papers working on a range of directions in a field, Main Path Analysis gives us a high-level view on the most important and most critical publications at any time. However, ”most important” is a challenging term. Many publications may be critical and significant at the time, but over a period of decades, research interests died down and contemporary researchers may not be aware of them. Are those publications still important? Einstein’s papers are influential partially because they are still being actively engaged with today. Main Path Analysis allows us to define the most important papers based on either their influence on and role in developing current trends or their central position in synthesizing earlier research at the time. In this chapter, I rely on the default use of Main Path Analysis to reveal key papers in the history of PET that ultimately contributed to the current state of development of PET.
To supplement Main Path Analysis, I also rely on Co-Citation Analysis. Co-Citation Analysis is a way to understand the relationship between a pair of publications. The general idea is that the more two publications share references, the tighter their content is coupled with each other (Small 1973). While it is difficult to interpret why two publications share many references, if two bodies of literature have substantial similarities in references, it is reasonable to assume that they are literature on similar or the same projects.

In order to collect the requisite data and to perform the Main Path Analysis, Co-Citation Analysis, and other bibliometric analyses, I relied on a series of steps to pull data from online databases. The steps are available for replication in the Appendix.

### 6.3 Adopting and Surviving New Technology

While the data collected and processed, I use this chapter to demonstrate the two-engine model of PET development based on bibliometric analysis. By “the two-engine model,” I mean that chronologically, and through Main Path Analysis, the field of PET research moves forward through identifiable phases of changing scientific topics. The field of PET may focus on one topic at a time, and move onto the next topic in five or ten years. However, at the same time, the research community is not primarily organized along research topics, but rather through researchers’ choice of technologies.

Picking up from the last chapter, around the end of the 1990s, researchers were not aware that PET research was coming to a turning point: In the late 1980s and early 1990s, a new and arguably better technology – functional Magnetic Resonance Imaging (fMRI) – was emerging and would eventually take over PET’s role as the leading technology to image the human brain. The growth of fMRI and its eventual taking over of PET in its share of functional imaging research is evident in Figure 6.1. Research related to fMRI started in the late 1980s. While research in fMRI was still maturing, research using PET had a visible bump around 1990, possibly because of Fox and Raichle’s seminal
paper on the use of PET for cognitive research (Fox et al. 1988). Then, as fMRI matured, research in fMRI developed in a pattern similar to exponential growth and ultimately overtake PET in human functional imaging research by around 1996. The popularity of fMRI is due to a number of reasons, including faster and higher resolution scans and not using radioactive isotopes to both reduce the radioactive exposure and the cost to maintain a cyclotron on site.

While fMRI quickly takes off and explodes, research in PET does not simply die down. Instead, researchers committed to the continued use of PET needed to identify new directions. It was similar
to when researchers invested in nuclear research around World War II had to pivot and find new uses of their nuclear research after the war. But in what way?

The Main Path Analysis nicely demonstrates the early years of PET research, its strength in the 1970s, and its new direction in the 1990s after the arrival of fMRI. I ran the Math Path Analysis using Batagelj’s algorithm (Batagelj 2003) on the database of bibliometric data on the more than 100,000 entries of publications collected through the steps listed in the Appendix. The algorithm generated the following graph:

Here, in the graph, the top denotes the most recent developments. The bottom denotes earlier developments. The arrow goes from top to bottom, or from recent research projects to earlier research projects, to demonstrate that the algorithm evaluates the significance of research projects and papers primarily based on their impact on current research projects. The algorithm begins with the most recent and important projects and trace back the pivotal projects and papers that contributed to the current landscape.

The Main Path Analysis results visually demonstrate the trajectory of the PET field from its disorganized early years to its pivoting in the 1990s. In its early years, researchers were using early versions of PET on a variety of research directions: studying how PET could be used to perform any measurement, studying how PET could be used to study cerebral metabolism of glucose, or studying blood volume change in the brain (all green nodes). The early years of PET cover the gap between the 1950s and 1970s, the part of history between the MGH years and the WUSTL years covered in the dissertation. Just as I have noted in Chapter 5, in the 1960s and 1970s, the research on PET was very disjointed as researchers were simply looking for an application of the novel technology that would stick in the academic or clinical research community. Such efforts were not particularly successful, but the diverse range of interests collectively attracted enough funding to support the existence of the nascent technology against more mature ones such as X-ray CT and SPECT.
As time goes on, the convergence of those directions in the 1970s formed the technical and theoretical basis of human functional PET: human functional PET was born out of the technical capacity to image brain blood volume and glucose metabolism and the theoretical understanding that they are correlated with cognitive activities. By combining both the technology and the theoretical understanding, researchers developed the PET technology for human functioning as we know it: in the early 1980s, encouraged by the researchers at WUSTL, researchers using PET concentrated...
on the relationship between physiological changes in the brain and the cognitive changes (the pink nodes).

But the research trajectory radically changed with the arrival of fMRI. fMRI gradually became the preferred choice of human functional imaging, especially in the 90s, replacing PET from the research field. Fortunately, in the 1980s, instead of just studying cognition, researchers also started to explore the clinical value of brain imaging (blue nodes). Those efforts to use PET in clinical settings in a way prepared PET’s pivoting when fMRI arrived. Here, I must note the particularity of the Main Path Analysis algorithm. In the 1980s, Fox and Raichle came up with their seminal paper on using PET for human functional neuroimaging (Fox et al. 1988). But such a paper did not show up in the Main Path Analysis because the paper came too late: fMRI quickly dominated the field of human functional neuroimaging and PET research ended up pivoting to various clinical projects. Because I choose to use the Main Path Analysis to assess significance of papers based on their influence on current research agenda, and because Fox and Raichle’s seminal paper had limited influence on the current, clinical uses of PET, the paper does not show up in the analysis. Instead, research into clinical uses of PET, although not a major influence in the 1980s, becomes significant in Main Path Analysis because current research critically depends on such endeavors.

In the 1990s, the PET research field mostly moved on from human functional imaging research. Continuing from clinical efforts in the 1980s, the researchers using PET bifurcated: some chose to focus on the study of cancer and others chose to focus on cerebrovascular disease (orange and green nodes). In those fields, because researchers using PET can use specific isotopes to obtain information regarding cancer and cerebrovascular disease in ways that MRI could not, research using PET was able to maintain its old level of activity, as indicted in Figure 6.1.

In short, the Main Path Analysis demonstrates that shifts in major research topics and projects are backbones and drivers of the history of PET research. The ideas of using PET to research all kinds of
bodily functions gave way to research on cognition, which in turn gave way to clinical applications on cancer and others.

However, Co-Citation Analysis reveals a very different picture: researchers and research projects primarily associate with each other along their choices of the technology, instead of their shared research interests. In Figure 6.3, the visual and physical distance between nodes is the distance of their co-citational strength: the closer they appear, the more citations they share. Although both PET and fMRI are used for functional neuroimaging, the Co-Citation map of the entire body of literature created in this dissertation demonstrates a clear separation between the PET research (on the lower left) and fMRI research (on the upper right). Even within each group, researchers arrange themselves according to their choice of technology. The brown and green clusters on the upper right largely represent fMRI research in the body of literature. The brown cluster consists primarily of background

Figure 6.3: Clusters of Publications
theories and methodologies of fMRI, such as the discovery of neurobiological basis that enables the technology of fMRI (Fox and Raichle 1986). The green cluster consists primarily of the fMRI technology itself, such as its early, seminal papers that propose its key technical mechanisms (Ogawa et al. 1990; Kwong et al. 1992).

The clusters in the lower left largely represent PET research: The purple cluster consists primarily of high impact background theory papers, including foundational statistical methods (Friston et al. 1991), theoretical descriptions of the attention system (Posner and Petersen 1990), and those of the visuomotor system (GRAFTON et al. 1992), among others. The blue cluster consists primarily of specific research projects in neuroscience, such as studying the neural basis of memory (Paulesu, Frith, and Frackowiak 1993; Ungerleider 1995; Rugg et al. 1996). The orange cluster consists primarily of specific and practical methodology papers, such as the assessment of handedness (Oldfield 1971). The yellow cluster consists of publications mostly on neurobiology and neuroanatomy, such as papers on the cingulate cortex (Vogt, Finch, and Olson 1992) and emotion’s neuroanatomical basis (Lane et al. 1997). The red cluster consists of publications mostly on clinical research, such as those on Parkinson’s (Eidelberg et al. 1994) and Alzheimer’s (McKhann et al. 1984). The lower right light blue cluster consists of other related brain imaging technologies, primarily on FDG-PET, or the use of glucose metabolism tracers to study brain function (Theodore, Fishbein, and Dubinsky 1988; Gaillard et al. 1995).

The Co-Citation Analysis in Figure 6.3 demonstrates that PET and fMRI are reasonably separate technical and theoretical communities, although researchers use both technologies on human functional neuroimaging. And even within each subcommunity, the choice of specific technology impacts the structure of researcher subcommunities. The clusters demonstrate that the PET community and the fMRI community are much tighter connected within their own respective community instead of across the aisle. In a way, the community separation between PET and fMRI can be explained by the difference in theoretical knowledge and technical expertise needed to operate PET and fMRI. PET
relies on the theoretical principles that increased cognitive activities are associated with increased cerebral metabolism and blood flow, while fMRI relies on the theoretical principles that during increased cognitive activities, the cerebral oxygenation level is decoupled from blood flow (Fox and Raichle 1986). PET relies on technical mechanisms to capture escaping radiation from the brain, especially through the use of coincidence detection to capture and count pairs of gamma rays. fMRI relies on detecting magnetic changes in the blood oxygenation level under a high magnetic field (Ogawa et al. 1990; Kwong et al. 1992). Due to the theoretical and technical differences, despite both technologies are used in human functional neuroimaging, the respective research communities and publications have relatively sparse connections across two technologies.

But what about sub-communities with researchers who primarily use PET? It seems that, in Figure 6.3 communities are organized by their use of different isotopes, as the study of cerebral activities, clinical research, and FDG all have their own choice of isotopes. Why are researchers organized around specific technologies than their research projects? In the next subsection, I rely on insights from the evolutionary metaphor to suggest that the community structure is likely and partially the result of researchers searching for the competitive advantage of PET technology.

### 6.4 Modularity and Community Ecology

In the Introduction and Chapter 2, I have already briefly discussed modularity and community ecology as key elements in the evolutionary metaphor I use in the dissertation. In this chapter, I rely on modularity and community ecology, to shed more light on the “two-engine” model of scientific instrumentation as reflected in the history of PET. In particular, I argue that the modularity of PET not only allowed PET to survive in its early and immature years, but also contributed to community structure in which researchers organize themselves around technological choices, instead of just around scientific ones.
Modularity, in essence, is a feature of complex systems that they can be broken down into functional sub-systems and components. Following Simon’s work (Simon 1962, p. 468), contemporary biologists propose the following four characteristics for modularity (Callebaut and Rasskin-Gutman 2005, p. xv):

• Hierarchical System: Modularity takes place in a system which can be decomposed into sub-systems.

• Near-decomposability: Systems and their sub-systems can be reasonably cleanly decomposed functionally and structurally.

• Quasi-functional-independence: “Modules” – functional and structural components – should be more or less self-contained functional and structural units that can perform their functions on their own. It is understood that modules do need input and output from the outside. Sometimes modules are dependent on other modules in other ways. Therefore, independence is always a degree and modules most often are only quasi-, instead of fully independent.

• Recurrence: Biologists can often identify similar functional units across a very wide spectrum of animals (Müller and Newman 2003).

Contemporary research has suggested that modularity is fundamental, if not absolutely necessary to evolution by natural selection. Lewontin, for example, states that modularity is a necessary condition for adaptation (R. C. Lewontin 1978, p. 229):

Continuity means that small changes in characteristic must result in only small changes in ecological relations; a very slight change in fin shape cannot cause a dramatic change in sexual recognition or make the organism suddenly attractive to new predators. Quasi-independence means that there is a great variety of alternative paths by which a given characteristic may change, so that some of them will allow selection to act on the characteristic without altering other characteristics of the organism in a countervailing
fashion; pleiotropic and allometric relations must be changeable. Continuity and quasi-
independence are the most fundamental characteristics of the evolutionary process. Without them organisms as we know them could not exist because adaptive evolution would have been impossible.

Here, Lewontin says that it is a necessary condition (“without them [...] adaptive evolution would have been impossible”) that specific functional modules could be selected on and evolve without much influence on other modules.

Modularity is fundamental to evolution by natural selection partially because modularity enhances evolvability. Evolvability here refers to the extent to which a population can produce better fitting, more adaptive offspring in any environment. “Fitness” and “adaptation” are both loaded, and sometimes controversial terms in evolutionary biology. Fortunately, we can understand the usefulness of modularity without jumping down the rabbit hole of theoretical and terminological debates surrounding “fitness” and “adaptation.” Many theories of why and how modularity enhances evolvability exist. Here, I am adopting a version of John H. Holland’s theory (Holland 1992). Assume, for a moment, evolution by natural selection is nothing more than searching for better solutions in a massive solution space. Some fundamental properties of an evolving population should include:

- It should be able to store the traversed path and lessons learned along the road
- It should be able to access the learned knowledge
- It should be able to exploit the knowledge while searching for better solutions

It quickly becomes clear that a non-modular population - a system that cannot go through partial upgrades – cannot easily store or access the traversed path. For a non-modular population, once an offspring mutates, it goes through a wholesale change and loses properties or functions that worked for its parents. A modular population, however, can meet all three requirements. Assume we have an organism with 8 functional modules and any single offspring only modifies one module.
Evolution by natural selection can slowly retouch functional units without interfering with other units. Evolution by natural selection can make a wing from a leg without disrupting the heart, the lung, or the brain. Evolution by natural selection can search for solutions in a massive solution space by modifying one module at a time. Doing so helps organisms store, access, and exploit the information they have learned throughout the evolutionary history.

The evolvability point seems to apply perfectly well in engineering in general and particularly with scientific instrumentation. Many researchers have already observed and commented on the similarity between evolution by natural selection and engineering. The following is an observation to connect a tinkering engineer with the natural selection process by François Jacob, a Nobel prize winner: (Jacob 1977, pp. 1163-1164)

[Natural selection] works like a tinkerer – a tinkerer who does not know exactly what he is going to produce but uses whatever he finds around him whether it be pieces of string, fragments of wood, or old cardboards; in short it works like a tinkerer who uses everything at his disposal to produce some kind of workable object. For the engineer, the realization of his task depends on his having the raw materials and the tools that exactly fit his project. The tinkerer, in contrast, always manages with odds and ends. What he ultimately produces is generally related to no special project, and it results from a series of contingent events, of all the opportunities he had to enrich his stock with leftovers.

Often, without any well-defined long-term project, the tinkerer gives his materials unexpected functions to produce a new object. From an old bicycle wheel, he makes a roulette; from a broken chair the cabinet of a radio. Similarity evolution makes a wing from a leg or a part of an ear from a piece of jaw.
Evolution behaves like a tinkerer who, during eons upon eons, would slowly modify his work, unceasingly retouching it, cutting here, lengthening there, seizing the opportunities to adapt it progressively to its new use.

Jacob tells us something interesting about evolution by natural selection that we learn by analogy with engineering. But interestingly, he misses the fact that engineering can learn something from biology too, especially that the evolution of technology is in some instances more like natural selection. In particular, engineers almost never have a completely preconceived plan in which the engineer “foresees the product of his efforts” (Jacob 1977, p. 1163). Nor do engineers have “materials specially prepared to that end and machines designed solely for that task” (p. 1163). Nor do engineers often “approach the level of perfection made possible by the technology of the time” (p. 1163). Instead, as what I have discussed in previous chapters, engineers are more like what he describes as tinkerers: who make do with materials in hand, who move forward because they see opportunities with the materials, not because they have a preconceived plan waiting for materials to fall into place. But his broad point stands: the process of evolution by natural selection, by using and re-using existing parts and modules, is extremely similar to the common engineering process. Engineered products, including scientific instruments, are indeed modular too in the sense that we can retain the working modules and replace the non-working modules.

Let us return to the microscope example I used in the Introduction (Hacking 1981). In the seventeenth century, the microscope suffered from bad optical aberrations and produced blurry images. But improvements in optics and craftsmanship created better lens with tolerable optical aberrations. Scientists and amateurs did not need to completely re-design the microscope to take advantage of the new lens. They just had to replace the old lens with the better lens and they improved on the microscope and received scientific adoption.

Just like the importance of modularity for the microscope, I argue that modularity is a key success factor for Positron Emission Tomography. Because the instrument is modular, researchers could
make progress with technological developments without a deeper theoretical understanding of cerebral processes. Because the instrument is modular, researchers were able to swiftly adopt new developments in nuclear research into existing medical instrument designs. Also because the instrument is modular, researchers were able to come up with instrument designs that may potentially work in a wide variety of medical disciplines and quickly identify potential venues for adoption.

I am here to list a few areas where researchers relied on modularity to improve on their design as I have covered in the previous chapters:

- The choice of a good crystal: between the 30s and 50s, researchers spent decades on searching for better crystals. They found a crystal that worked exceptionally well with high-energy gamma rays (the NaI(Tl) crystal), combined it with the photomultiplier and then had a scintillator to work with gamma rays. Developments in finding better crystals did not require significant redesigning of the system or understanding of cerebral biology. The choice of NaI(Tl) lasted all the way throughout the 80s for a wide range of medical uses.

- The choice of a good radioactive isotope: the team at MGH initially realized that Potassium-42 was unsafe to the human being. They waited for chemistry and biology to catch up and offer better radioactive isotopes. Within just a few years, they found Copper-64 which was much safer to the human being. Later teams at WUSTL and other research institutions used Fluoride-18, Oxygen-15, and other radioactive isotopes without massively redesigning the PET technology.

- Coincidence detection: the move to coincidence detection greatly improved signal-to-noise ratio and sensitivity of the team’s detection system. To adopt coincidence detection, researchers did not redesign their system. Instead, they took off the collimator in front of the crystal (think of taking off the hood in front of the lens in a camera), and paired two cameras in opposite directions.
• Multiplying the number of detectors: after the MGH team realized that merely a pair of detectors were not efficient in capturing radioactivity that goes in all directions, they decided to increase the number of detectors. In the end, they came up with a design of 127 pairs of crystals.

• Adopting PET for human functional neuroimaging: the WUSTL relied mostly on existing modules developed at other research institutions, improved on them, and adopted them to scan the brain fast. Those modules were not developed for or based on brain scanning.

In this chapter, I further the observation that modularity is not merely important for PET researchers to pivot and identify ways to generate academic interest but also critical in shaping the structure of PET community. In particular, I rely on previous chapters’ discussion on the choice between using oxygen isotopes versus using fluoride isotopes in PET research. The modular PET technology can work with both oxygen isotopes and fluoride isotopes with only minor changes. But the isotopes have different uses: oxygen isotopes are extremely useful in studying cognitive processes, while fluoride isotopes are useful in studying metabolism, including cerebral metabolism and cancer. Because switching between the two isotopes and two directions is relatively easy in terms of PET technology, the modularity feature of PET allowed researchers to appeal to and, as recorded in Chapter 5, even create academic trends.

At the same time, the choice between oxygen and fluoride isotopes likely created two PET communities. While PET could work with both and more isotopes, manufacturing and procuring oxygen and fluoride isotopes are entirely different processes. Because the oxygen and fluoride isotopes have different half-life and equipment requirements, they contributed to the segregation of PET communities. Fluoride isotopes have a long half-life and can be manufactured by nuclear reactors offsite. So researchers can purchase fluoride isotopes from outside suppliers and use them without committing to installing and maintaining a nuclear reactor. Oxygen isotopes have a very short half-life and must be manufactured with an on-site cyclotron. The practical needs and the requisite
technical knowledge to operate two different radioactive isotopes likely contributed to the specific communal structure: to rely on the metaphor of community ecology, researchers were consuming “resources.” In the case of PET, those resources include institutional resources to build and maintain research equipment (including a multimillion-dollar cyclotron) and professional resources to acquire knowledge regarding specific isotopes and their related use in medicine. Researchers with limited professional resources may need to choose to focus on one set of isotopes over another. And such a choice may be partially dictated by institutional resources: more well-funded universities and medical schools may be able to afford to install and maintain a cyclotron, while other universities and medical school could only afford to order fluoride isotopes from outside sources.

The segmented sub-communities in PET research perhaps also contributed to the growth and appeal of PET research. The availability of both the oxygen and fluoride options allowed the low institutional resource research community to work on metabolism-related research using PET and produce exciting research results. The high institutional resource community can work on cognitive processing and find their own direction of publication.

The modularity nature of PET allowed it to appeal to broader but segmented communities without substantial technical changes. At the same time, activities in multiple research sub-communities allowed PET to survive when fMRI started to replace PET in human functional neuroimaging. Let us review the Main Path Analysis timeline again:

- The adoption of earlier research on using oxygen isotopes to measure cerebral blood flow and using fluoride isotopes to measure cerebral metabolism
- The concentration of instrumentation efforts on human functional neuroimaging
- The re-use of the PET technology on cancer, cerebrovascular disease, and other clinical uses

In all of those cases, researchers in PET could work in segmented sub-communities, develop their own research directions, and dynamically respond to the changing landscape of academic and
When PET is no longer appeal in the academic research of human functional neuroimaging, its clinical research community started to take over. In the entirety of this process, researchers on PET did not need to completely re-design PET technology. Nor did they need to wait for theoretical developments to mature. Instead, they could iteratively replace modules to improve their performance or change the direction of their PET technologies. Modularity of scientific instrumentation allowed researchers on PET to learn from paths already traveled by earlier generations, better adapt to changing academic needs and landscape, and find the ways of technical adoption, acceptance, and survival.

As a contrast, because fMRI is a wholesale redesign relative to PET, fMRI cannot preserve many of the useful features of PET. PET, by using many isotopes, can study the chemical process involved in cognition but fMRI is primarily dependent on measuring oxygenation level changes during cognition. Therefore, there are many efforts to combine fMRI and PET, instead of ditching PET, in both clinical research and neuroscientific research. One such example is the study of pain experience. By combining fMRI and PET, researchers can understand both the brain oxygenation changes and the opioid-related chemical experience during pain experience (Wey et al. 2014). The co-existence of fMRI and PET demonstrates that complete wholesale design is often not conducive to preserving what researchers have learned and have worked out in history.

### 6.5 Conclusion

In this chapter, I used bibliometric analysis to illustrate how the PET field changed after research at MGH in the 1950s and provided a dynamic, causal account of technological development and innovation dissemination through focusing on the modular feature of scientific instruments.

The bibliometric method is a computer-assisted method to extract information from a large body of literature. It enabled me to comb through more than 100,000 academic publications on PET and fMRI and identify relationships between those publications. The results clearly illustrated the
path and patterns of PET development between the 1970s and the 1990s. They also demonstrated the relationship between PET and fMRI: they largely occupied different communities because of distance in their theoretical principles and technical mechanisms.

Perhaps the most important insight from the bibliometric analysis is the key role of modularity and community ecology in scientific instrumentation. Modularity – a feature of complex systems that they can be broken down into quasi-independent structural and functional modules– allow researchers to develop instruments one piece and one step at a time. They can preserve what works at a time and improve on other elements or adjust to later changes in demands. They can develop modules without a full understanding of the overarching social demands or theoretical understanding at a later time. They can also easily pivot to find new uses. In this chapter, modularity is particularly important in helping to explain how PET survived its earlier years, pivoted in its later years after fMRI took over the field of human functional neuroimaging, and most importantly the “two-engine” structure of scientific instrumentation. While scientific projects and topics drove the development of PET technology and research, its modular feature allowed it to appeal to different technology-driven and technology-defined sub-communities. The existence of those sub-communities not only broadened the appeal of PET, but also gave PET research enough flexibility when new challenges – in the form of fMRI – came around the corner. PET is still alive and well, partially because of modularity.
Chapter 7

Conclusion

In the early 2000s, Chang and Leonelli published two papers on the ontology of the infrared light (Chang and Leonelli 2005a, 2005b). In the two papers, they recount and examine a scientific debate in the nineteenth century: whether light and heat were identical. They focus on how scientific practices and evidence at the time chimed in on two incompatible theories: the pluralistic theory of radiation versus the unified theory. The pluralistic theory stated that the radiation that produced heat was not the same radiations that produced light (Chang and Leonelli 2005a, p. 489). The unified theory of radiation stated that the same kind of radiation produced both light and heat. Chang and Leonelli explicitly frame their history as an episode of theory choice (p. 490). They suggest that, based on their review of the evidence at the time, there was no prevailing reason to prefer the unified theory over the pluralistic theory (p. 505). In fact, many anomalies existed for the unified theory. The pluralistic theory was perhaps just as good as the unified theory at the time.

Their history, at the same time, is a history of technology-driven scientific progress. Researchers in the nineteenth century did not have preconceived theories of radiation first and then turned around to discover evidence to justify or refute their theories. Instead, the debate began with new instruments to separate the visible light into bands of spectrum and to detect trace amount of heat. For example,
the following is a piece of history recounted by Chang and Leonelli: (Chang and Leonelli 2005a, pp. 479-480)

Herschel had stumbled upon the effect in question while trying to observe the sun with a telescope, which required strong eye-protecting filters. While testing out various filters, Herschel noticed that the light coming through differently colored filters seemed to carry different amounts of heat. He then made more precise observations by placing thermometers in various parts of the solar spectrum projected through a prism. Noting that the heating power increased as he moved toward the red end of the visible spectrum, Herschel tested whether it might continue into the dark space beyond the red, and indeed detected a great heating effect. This started him on a course of investigations, whose sensational results he reported in four successive papers in the *Philosophical Transactions of the Royal Society of London*.

In this case, Herschel, a prominent scientist of the time and the Court Astronomer for George III of the United Kingdom, used colored filters to separate a specific spectrum (signal) from other bands of spectrum, reduce the noise (undesirable bands of spectrum) in the observation, and got to study visible light spectrum band by band more precisely. Herschel was not the only one intent on developing new instruments and pursuing better exclusion of noise. Macedonio Melloni, another establish physicist of the nineteenth century, devised even more advanced techniques to study light. He came up with a way to cleanly separate ultraviolet and detect its extremely faint level of heat (p. 469). He also came up with ways to further reduce noise in his measurement, by ensuring rays of different wavelength did not get mixed together, and to ensure the heat detector was narrow enough to measure the heat only from a very specific band of light spectrum, etc. (p. 469).

The history of research in infrared light shares many similarities with main themes covered in this dissertation. Similar to the history of PET, the history of infrared light offers many interesting lessons in the development of scientific theory, theory choice, evidential relationships, etc. At the
same time, instruments play a central and indispensable role in moving science forward. While the former – progress on scientific theory – commonly receives significant attention from philosophers of science, the latter – progress on instruments and their role in science – is often pushed to the background. I rely on the history of PET in this dissertation to supplement the philosophical focus on theory with a focus on instruments.

In particular, in this dissertation, I have focused on the relationship between engineering and epistemic norms and the development of technology. I have listed a number of norms, such as increasing signal-to-noise ratio, increasing sensitivity, and increasing resolution which are important in the history of PET and other detection technologies. Those norms are distinct from other norms, such as increasing accuracy and increasing reliability, at least in three ways:

- They are instrument-centric: those norms are specifically about improving features and functions of causal mechanisms
- They are diagnostic: those norms evaluate causal mechanisms and offer specific causal means to improve
- They address particular philosophical and scientific problems: inferring from data to phenomena through the mediation of instruments

Those norms influence and inform the development of technology similar to an evolution by natural selection process. In a way, those norms serve as “selective forces,” to which instruments and researchers not only respond, but also shape. Researchers can reorganize modules of their instruments and incrementally improve on separate modules to better appeal to those norms. They can also change the social context and the relative priority between those norms – such as by prioritizing temporal resolution. Among the factors in the social context is the excitement generated regarding scientific inquiries. For the researchers at Washington University in St. Louis, they managed to convince the academia the significance of imaging cerebral activities in vivo and in real time and
generated many novel scientific questions and answers in the process of developing their new instruments and modifying the competitive landscape.

The engineering and epistemic norms are important also because they “anchor” the epistemic value of data instruments generate. In many many cases, researchers have no direct access to their target phenomena and they cannot “directly” evaluate if their data is a good reflection of their target phenomena. In practice, in the case of PET and other instruments, researchers treat data as evidentially valuable because they are generated by good instruments. Instruments are good precisely because they fulfill the norms mentioned in this dissertation.

In the conclusion, I move beyond the history I have already told. Instead, I dive into other episodes of history to emphasize that the lessons I have learned from the history of PET are not unique to a single detection instrument.

The first example is of course the history of infrared light by Chang and Leonelli. In that episode of history, researchers developed tools to separate a highly specific band of spectrum from the full spectrum of light. They developed tools to detect heat generated by a narrow band of light from the entirety of heat in the environment. In the process, they developed new technologies to pursue after better signal-to-noise ratio. In other words, increasing signal-to-noise ratio was a prominent influence on scientific instrumentation and contributed to the later understanding of the relationship between light and heat through the study of infrared light as recounted by Chang and Leonelli.

The second example comes from Franklin on his recount of research on microwave radiation from Jupiter. In this episode of history, researchers intended to confirm if some of their received microwave radiation from Jupiter originated as synchrotron radiation (Allan Franklin 1989, p. 172). Their challenge was a challenge of separating and recovering information. Jupiter is known to be a very powerful source of radiation – emitting all kinds of radiation and many of them at very high energy levels. Synchrotron radiation is one among many forms of radiation out of Jupiter. It has a broad spectrum, which covers all the way from microwaves to high X-rays. It also has very unique
characteristics (e.g., it is polarized). Two of the challenges in the aforementioned discovery were to 1) distinguish synchrotron radiation from other sources of radiation from Jupiter and 2) identify synchrotron radiation as the source of the received microwave radiation.

When Franklin summarizes the history, this episode of history is primarily about theory-choice:

A somewhat different illustration of the complex interaction of observation and theoretical explanation is provided by the discovery of synchrotron radiation from Jupiter, or, to put it more accurately, the discovery that the microwave radiation emitted from Jupiter was due to synchrotron radiation. In 1959, Sloanaker reported measurements of the intensity of radiation from Jupiter at a wavelength of about 10 cm that gave temperatures, computed from a black-body model, that ranged from 300°K to 1,010°K, with a mean of 640 ± 85°K. Those temperatures were inconsistent with temperatures obtained from earlier measurements at 3 cm and in the infrared region. An additional problem was that the considerable scatter in the data allowed a variable component in the radiation. At the same time, Drake and Hvatum reported measurements at 22 cm and 68 cm that required temperatures of 3,000 °K and 70,000 °K, respectively, which they regarded as too high to be plausible. They combined their results with others and concluded that Jupiter was emitting nonthermal radiation. They proposed “that the radiation originates as synchrotron radiation from relativistic particles trapped in the Jovian magnetic field, a situation similar to the terrestrial Van Allen belts. A Jovian field of 5 gauss and a total number of particles 106 times greater than the terrestrial system will suffice to explain the observations.” There is at least a hint that the failure of the data to be consistent with the black-body model, combined with the variation and scatter in the data themselves, cast some doubt on the measurements.
Franklin focuses on how background models can validate or cast doubts on measurements and how measurements may adjudicate between competing models. Franklin summarizes the history as a classic theory-choice story.

But the original papers in the debate suggest that new and improved detection instruments enabled and moved forward the debate. For example, researchers relied on new detection instruments that were both high sensitivity (“Wild and Sheridan’s direction-finding technique, which incorporates phase switching in a swept-frequency interferometer, provides a useful increase in the sensitivity for our dynamic spectrograph.” (Warwick 1961, p. 39)), and high signal-to-noise ratio (“With a gain of about 9 db. over an isotropic radiator, the radio source Cassiopeia A stands out clearly above the fluctuations of total power.” (p. 39) (“a useful ...” and “stands out..” italics are mine. “Cassiopeia A” italics is original.)) Here, Warwick et al. started their paper on synchrotron radiation detection with an introduction of their technical improvement. They explicitly mentioned measures to increase the sensitivity and signal-to-noise ratio of their detection instrument.

In both Chang and Leonelli and Franklin, they focus their discussion on the development of scientific theories, one regarding the relationship between light and heat, and the other regarding the origin of microwave radiation. But in both episodes of history, the development of instruments, along the norms of increasing signal-to-noise ratio and increasing sensitivity, moved the science forward and provided the needed evidence for the respective scientific debate.

With the historical episode of PET, as well as those of light and heat and microwave radiation and many others, I want to showcase the significance of instruments and the associated norms in making scientific progress. I offer not just examples, but also causal mechanisms – relying on an evolutionary metaphor – on how researchers develop their new instruments and how new instruments may influence scientific inquiries. I hope my dissertation is merely the first step toward a shift and recentering of instruments in philosophy of science and philosophy of experimentation. Scientific theories do not live in an instrument-less vacuum.
References


Bagehot, Walter. 1872. Physics and Politics Or, Thoughts on the Application of the Principles of” natural Selection” and” inheritance” to Political Society. London: Henry King.


206


Charpa, Ulrich, and Ute Deichmann. 2007. *Jews and Sciences in German Contexts: Case Studies from the 19th and 20th Centuries*. Google-Books-ID: 00dDXnUck9QC. Mohr Siebeck.


208


Hooke, Robert, and Jo. Martyn and Ja. Allestry. 1665. *Micrographia, or, Some physiological descriptions of minute bodies made by magnifying glasses?with observations and inquiries thereupon/by R. Hooke ...* In collaboration with Missouri Botanical Garden. London : Printed by Jo. Martyn / Ja. Allestry, printers to the Royal Society ...


218


Appendix A

Step by Step Guide to Replicate the Data Analysis

In the Appendix, I list the steps to replicate the data collection process for the dataset used in Chapter 6 of this dissertation. All the scripts used in the Appendix are available from https://github.com/rainrdx/DigitalHumanities. In the data collection process, I have relied on access to two subscription-based databases: Web of Science and Scopus. Web of Science is owned by Thompson Reuters, which operates Science Citation Index (SCI), and offers metadata database services. Scopus is owned by Elsevier, and offers a similar but non-overlapping coverage of metadata of publications. Insofar as I am aware, the two databases do not have free access options.

The first challenge I am to address is the challenge of compiling a list of key words that capture the literature of PET and fMRI. But which “features,” or database search keywords identify such a dataset? Both the technology, PET, and the field, neuroscience, can be referred to in many ways. PET has a limited number of variations, such as Positron Camera, or Positron Emission Transaxial Tomograph. It is a small enough list to exhaustively enumerate. But researchers do not always announce that they are doing neuroscience. I decided to identify a useful but not necessarily
representative body of literature on using PET in neuroscience through review articles, extract common key words shared by those review articles, and then use those key words to search in both databases.

Hence, I organize my data acquisition steps as follows:

1. List major review articles in the field, such as:
   
   (a) Positron emission tomography: human brain function and biochemistry, Phelps and Mazziotta, 1985
   
   (b) Behind the scenes of functional brain imaging: A historical and physiological perspective, Raichel, 1998
   
   (c) PET: The Merging of Biology and Imaging into Molecular Imaging, Phelps, 2000
   
   (d) A Tabulated Summary of the FDG PET Literature, Gambhir et al., 2001
   
2. Download all the references listed in the select review articles that fulfill the following research criteria:
   
   (a) Is PET (instead of, say, fMRI)
   
   (b) Is about the brain (instead of other organs, full-body, animals, or plants)
   
   (c) Is a functional study of the brain (instead of stroke studies, tumor studies, etc.) (We are doing so because we are only interested in the functional study of the brain.)
   
3. Save all articles in folders by year (e.g. articles published in 1997 are saved in a folder named “1997”)
   
4. Use Adobe’s own OCR tools to extract the texts from PDFs and save them as text files
   
   (a) Open Adobe Acrobat Pro DC
(b) View -> Tools -> Enhance Scans -> Open -> Recognize Text -> in Multiple Files -> Add Folders -> Add “OCR” suffix to the OCR’d files.

5. Use pdftotext to extract the text files from OCR’d files

(a) Download pdftotext from http://www.foolabs.com/xpdf/download.html

(b) In the folder with all the OCR’d files, run

    for /r %i in (*OCR.pdf) do "pdftotext.exe" -layout "%i"

6. Save the text files in R’s working folder and run the following code:

    install.packages("tm")
    install.packages("SnowballC")
    install.packages("fpc")
    install.packages("cluster")
    install.packages("biclust")
    library("tm")
    library("SnowballC")
    library("fpc")
    library("cluster")
    library("biclust")

    (Text=VCorpus(DirSource(directory = "/home/rdx/Neuroimaging\Research\ Group/Scientific\ Publications", pattern="txt$", recursive = TRUE), readerControl= list(language= "lat")))

    Text <- tm_map(Text, removePunctuation)
    Text <- tm_map(Text, content_transformer(tolower))
    Text <- tm_map(Text, removeNumbers)
    Text <- tm_map(Text, removeWords, stopwords("english"))
    Text <- tm_map(Text, PlainTextDocument)
7. Find out the major clusters

8. Choice of Database

(a) The Database for Network Analysis

In order to conduct network analysis, I need information on title, year, list of authors and work cited in each article. Originally, I planned on using Web of Science to collect publications in neuroimaging. But it lacks abstracts or citation information for articles more than 25 years old. Because we need articles that date back to the early 1970s and I find no easy way to locate neuroimaging articles based on titles alone, I switched to Scopus. Scopus is a competitor of Web of Science offered by Elsevier. It indexes a significantly smaller number of journals than Web of Science, but it offers complete abstracts and citation information. In other words, I can locate publications in neuroimaging with the extensive abstract indexed way back decades ago and use the citation information to conduct network analysis. So I decide on Scopus.

(b) The Database for the Number of Publications

I decided to use Google Scholar, because it indexes the full-text of articles dating way back to the beginning of time. With the help of full-text search, it is much easier to
locate all neuroimaging publication. Unfortunately, although Google Scholar also has fabulous citation information, there is no way to export its data into machine readable form. So I can’t use Google Scholar for Network Analysis. (It is possible, however, if we have some funding, to write a program and run it for a few weeks on a server. But it will be costly given the amount of data Google Scholar has.)

9. Choice of Search Key Words

Clustering in R delivers two “clusters” of words that appear together very often: Cluster 1: “brain,” “cerebral,” “blood,” and “flow” and Cluster 2: “regions,” “activity,” and “changes.” I added “utilization” and “metabolism” to the list because they appear quite often as well. The following is the full search command: (TITLE-ABS-KEY(PET OR PETT OR fMRI OR MRI OR positron) AND TITLE-ABS-KEY(brain OR cerebral) AND TITLE-ABS-KEY((region AND activity AND change) OR (blood AND flow) OR utilization OR metabolism))

The command is explained as follows:

(a) TITLE-ABS-KEY means to search in titles, abstracts and keywords generated by the indexing service (here, Scopus).

(b) “PET OR PETT OR positron” limits our searches to PET, which is the topic of interest to us.

(c) Instead of searching for PET alone, I request articles that also contain words mentioned above, because we are interested in the application of PET, instead of its engineering details (which may be a topic for another day) or it’s deep down chemistry.

With the initial collection of data based, I then “snowball” the literature by identifying publications citing and cited by the aforementioned literature. The above process generated more than 100,000 non-duplicate entries of publications, across articles, book chapters, and books.