THE DISCOVERY-INVENTION CYCLE: BRIDGING THE BASIC/APPLIED DICHOTOMY

Venkatesh Narayanamurti, Tolu Odumosu & Lee Vinsel

February 2013
THE DISCOVERY-INVENTION CYCLE: BRIDGING THE BASIC / APPLIED DICHOTOMY

VENKATESH NARAYANAMURTI, TOLU ODUMOSU & LEE VINSEL
ACKNOWLEDGEMENTS

The authors would like to thank our colleague Laura Diaz Anadon for her valuable comments and suggestions to improve the quality of the paper.
ABSTRACT

Since the Second World War U.S. science policy has, in large measure, been driven by Vannevar Bush’s Science, The Endless Frontier. Bush’s separation of research into “basic” and “applied” domains has been enshrined as status quo in much of U.S. science and technology policy over the past seven decades. However, the relationship of science and technology research to economic and societal wellbeing requires a coherent national innovation policy to bridge this divide. Much of the debate over the categories of basic and applied centers on what the appropriate federal role is in innovation. Bush argued successfully that funding “basic” research was a necessary role for government, with the implication that “applied” research should be left to the auspices of markets except where necessary to the public interest (defense and health “applied research” for example). Nevertheless, the original distinction is an artificial one and this is perhaps the time that policy discussions are moved beyond a simple focus on the stated goals of research to more productive considerations of a holistic view of the research enterprise.

In this paper we hope to provide an alternative point of view. By examining both the evolution of the famous “linear model of innovation” -- which holds that scientific research precedes technological innovation -- and the problematic description of engineering being “applied science” we seek to challenge the existing dichotomies between basic / applied research, science and engineering, tracing how knowledge travels between different knowledge domains through a case study of a selected group of Nobel Prizes in physics.

Our conclusion is that the relevant viewpoint is the one which considers research involving long-time horizons and embraces the inherent bidirectional relationships between science and technology in the design of institutions and the allocation of resources. This conclusion makes it imperative that we discover new ways to bridge the arbitrary divide between basic and applied in our public policy debates, bringing all forms of research into deeper congress. Such a move can lead to both game-changing discoveries and inventions as well as to a more sustainable S&T ecosystem.
# TABLE OF CONTENTS

From Bush to Stokes: Efforts to Rationalize the Basic Applied Dichotomy .......... 1

Moving away from “Basic vs. Applied” toward “Invention and Discovery” ............ 2

The Information Age and Nobel Prizes in Physics ............................................. 4

Conclusions ............................................................................................................. 10

Implications for Research Policy ............................................................................ 11

References ............................................................................................................... 12
FROM BUSH TO STOKES: EFFORTS TO RATIONALIZE THE BASIC / APPLIED DICHOTOMY

Though some have claimed that the linear model arose directly from Vannevar Bush’s classic post-WWII work, Science, The Endless Frontier, the roots of the linear model are older. In some regards, the linear model builds on some long-standing ideas in Western culture (and other cultures as well) that favor the work of the “head” — or intellect — over that of the “hand”; including, in this case, technical expertise.¹

In this paper, we are more concerned with the aspects of the linear model that deal with research. While the steps required to introduce technologies into the marketplace successfully are important, of greater interest to us is the implied relationship between science and technology in the language of “basic” and “applied” research.

The question of whether understanding always precedes invention has long been a troubling one. For example, it is widely accepted that many technologies reached relatively advanced stages of development before detailed scientific explanations about how the technologies worked emerged. In one of the most famous examples, James Watt invented his steam engine before the laws of thermodynamics were postulated. In fact the science of thermodynamics owes a great deal to the steam engine. This undermines the background assumptions in the “basic” vs. “applied” category, and some initial steps toward a new vision for research practice have already been taken.

In 1997, Donald Stokes’ Pasteur’s Quadrant: Basic Science and Technological Innovation was published posthumously. In this work, Stokes argued that scientific efforts were best carried out in what he termed “Pasteur’s Quadrant,” that is, spaces that are motivated simultaneously by expanding understanding and by improving our abilities (technological, including medicine) to improve the world. Stokes’ primary contribution was in expanding the linear model into a two-dimensional plane that sought to integrate the idea of the unsullied quest for knowledge with the desire to solve a practical problem.

Stokes’ model comprises three quadrants, each exemplified by an historical figure in science and technology. The “pure basic research” quadrant, exemplified by Niels Bohr, represents the traditional view of scientific research as being inspired primarily by a desire to extend fundamental understanding; the “pure applied research” quadrant, exemplified by Thomas Edison, represents the classical inventor, driven to solve a practical problem; and the third, Pasteur’s Quadrant, exemplified by Louis Pasteur, represents the perfect mix of the two, inventor and scientist in one, expanding knowledge in the pursuit of practical

¹ Shapin, The Scientific Life; Stokes, Pasteur’s Quadrant.
problems. Stokes described this final quadrant as “use-inspired basic research”.

The publication of Stokes’ book was itself a moment of celebration and excitement. To think that that the science policy and academic communities would lay the vagaries of the linear model to rest once and for all. A blurb on the back of the book quotes U.S. Congressman George E. Brown, Jr.: “Stokes’ analysis will, one hopes, finally lay to rest the unhelpful separation between ‘basic’ and ‘applied’ research that has misinformed science policy for decades.” However, while clearing the ground for future research, Stokes often did not go far enough nor did his work result in sufficient change in how policy makers discuss and structure research. For, while Stokes noted how “often technology is the inspiration of science rather than the other way around,” his revised dynamic model does not recognize the full complexity of innovation, preferring to keep science and technology in their own paths that only mix in the shared agora of “use-inspired basic research.” It is also instructive to note that Stokes’ framework preserves the language of the linear model in the continued use of the terms basic and applied as descriptors of research.

We, however, seek to build an ontology that captures the complex interplay between the forces of innovation, an “integrated model of innovation,” as well as to recognize that radical innovation often only arises through the radical integration of science and technology.

**Moving away from “Basic vs. Applied” toward “Invention and Discovery”**

Our deepest reservation about Stokes’ model, however, is this: Stokes attempts to complicate the linear model by moving from a one-dimensional line traversing the end-points of basic and applied research to a Cartesian plane where they inhabit separate orthogonal axes. However the downside to Stokes’ Cartesian plane is that it preserves the language of basic and applied research. For example Pasteur’s quadrant is characterized as “use-inspired basic research.”

Perhaps more importantly, the efficacy and effectiveness of the research endeavor cannot be fully appreciated in the limited time frame captured by a singular attention to the motivations of the researchers in question. Motivations are important. Working toward finding a cure for cancer or to advance the frontiers of communications can be a powerful incentive, stimulating ground breaking research. However, motivations are only one aspect of the research process. To more completely capture the full arc of research it is important to consider a broader time scale than that implied by just considering the initial research motivations. Expanding the focus from research motivations to also include questions of how the research

---

2 One only need review the 2012 State of the Union (SOTU) address, or any of the last two SOTUs for examples of the persistence of the “basic” vs. “applied” frame. There are of course, exceptions to this observation. Two notable ones are the book by the National Academies, *Rising Above the Gathering Storm* and the report by The Royal Society, “The Scientific Century.”

3 Stokes, *Pasteur's Quadrant*, 84.
is taken up in the world, and how it integrates both science and technology allows us to escape the basic/applied binary. It is important to consider the future oriented aspects of research as well as the initial motivational aspect of research. Considering the implications of research in the long term requires an emphasis on visionary future technologies, taking into account the wellbeing of society, and not being content with a porous dichotomy between basic and applied research.

As such, we propose using the terms “invention” and “discovery” to describe the twin channels of research practice. For us, Invention is the “accumulation and creation of knowledge that results in a new tool, device or process that accomplishes a particular, specific purpose, while Discovery is the “creation of new knowledge and facts about the world.” Considering the phases of invention and discovery along with research motivations and institutional setting enables a much more holistic, long-term view of the research process. This allows us to examine the ways in which research generates innovation and leads to further research in a virtuous cycle.

Innovation is a complex, non-linear process. Still, straightforward and sufficiently realized representations like Stokes’ Pasteur’s quadrant are useful as analytical aids. To such an end, we propose the model of the “discovery-invention cycle” that will serve to illustrate the interconnectedness of the processes of invention and discovery, and the need for consideration of research effectiveness over longer time frames than is currently the case. Such a model would allow for a more reliable consideration of innovation through time. The model could also aid in discerning possible bottlenecks in the functioning of the cycle of innovation, indicating possible avenues for policy intervention.

To summarize, we advocate a shift from thinking about research in terms of motivations (basic vs. applied) alone to a broader consideration of the “mirror-image twins” of discovery and invention. We also believe that research should be evaluated not only by its initial motivations, but also by the effect it has in catalyzing other research and innovations. This sort of analysis requires taking the longer time view into perspective.

---

4 The question of how research is taken up in the world is an important one and should be considered separately from questions about research motivations. Motivation to undertake pure research into an esoteric topic could conceivably lead to inventions that might have wider applicability beyond the initial research question and vice-versa. Motivations are important, but the fruits of the research project are at times even more important than the initial impetus.

5 By innovation we mean “the process of generating original ideas and insights of value that have the capability of making a positive difference in the world.”

6 Considering research, not as a single isolated project, but as part of a specific trajectory over a long time horizon is critical to the notion of the discovery-invention cycle. It allows for contextualization and visioning of the future to shape the dimensions of the current research project.

7 See Layton for metaphor, “Mirror-Image Twins.”
To illustrate this idea, consider Figure 1 in which we trace the evolution of the current information and communication age. What can be said about the research that has enabled the recent explosion of information and communication technologies? How does our model enable a deeper understanding of the multiplicity of research directions that have shaped the current information era? To fully answer this question, it is necessary to examine research snapshots over time, paying attention to the development of knowledge and the twin processes of invention and discovery, tracing their interconnections through time. To our mind, the clearest place for selecting snapshots that illustrate the evolution of invention and discovery which enables The Information Age, is the Nobel Prize awards.

We have thus examined the physics Nobel Prizes from 1956, 1964, 1985, 1998, 2000, and 2009, awards all related to information technologies. In addition, by applying our framework to the “hard case” of physics Nobel Prizes, we seek to show that the discovery-invention cycle model is just as valid for the “Sciences” as it would be applied to “Technology.” We describe these kinds of clearly intersecting Nobels as a “family” of prizes in that they are all closely related. We have identified other such families whose innovation cycles can be clearly described and illustrated through time.

The birth of the current Information Age can be traced to the invention of the transistor. This work was recognized with the 1956 physics Nobel Prize awarded jointly to William Shockley, John Bardeen, and Walter Brattain “for their researches on semiconductors and their discovery of the transistor effect.” Building upon early work on the effect of electric fields on metal semiconductor junctions, the interdisciplinary Bell Labs’ team built a working bipolar-contact transistor and clearly demonstrated (discovered) the transistor effect. This work and successive refinements enabled a class of devices that successfully replaced electromechanical switches, allowing for successive generations of smaller, more efficient and more intricate circuits. While the Nobel was awarded for the discovery of the transistor effect, the team of Shockley, Bardeen, and Brattain had to invent the bipolar-contact transistor to demonstrate it. Their work was thus of a dual nature, encompassing both discovery and invention. The discovery of the transistor effect catalyzed a whole body of further research into semiconductor physics, increasing

---

8 By analyzing the processes of invention and discovery in physics Nobel prizes, long considered the most mathematical and analytically precise of the sciences, we are continuing the fine tradition of undermining the specious claim that science precedes technology.

9 For example, the many physics and chemistry Nobel prizes concerned with nuclear magnetic resonance and imaging.

10 Interestingly, in his lecture, Shockley discusses the usefulness of the classifying terms “pure,” “applied,” “fundamental,” “basic” as they are used to describe research in science and technology. For Shockley, these terms are too often used in a derogatory sense to elevate research that is driven by a motivation of “aesthetic satisfaction” over research driven by a desire to improve a process. Shockley, “Nobel Lecture Presented in Stockholm,” 345.
knowledge about this extremely important phenomenon. The invention of the bipolar contact transistor, led to a new class of devices that effectively replaced vacuum tubes and catalyzed further research into new kinds of semiconductor devices. The 1956 Nobel is therefore exemplary of a particular kind of knowledge making that affects both latter discoveries and latter inventions. We call this kind of research, _radical innovation_. In Figure 1 the 1956 prize is situated at the intersection of Invention and Discovery and it is from this prize that we begin to trace the Innovation Cycle for the prize family that describes critical moments in the information age.

The second prize in this family is the 1964 Nobel Prize, which was awarded jointly, one-half to Charles Townes and the other half to both Nicolay Basov and Aleksandr Prokhorov. The 1964 Nobel Prize was given for work that is fundamental to the modern communications age. Most of the global communications traffic is carried by transcontinental fiber optic networks, which utilize light as the signal carrier. Townes’ work on the stimulated emission of microwave radiation earned him his half of the Nobel. This experimental work showed that it was possible to build amplifier oscillators with low noise characteristics capable of spontaneous emission of microwaves with almost perfect amplification. The _maser_ effect, “microwave amplification by the stimulated emission of radiation,” was observed in his experiments. Later, Basov and Prokhorov, along with Townes, extended the maser effect to consideration of its application in the visible spectrum and thus the laser was invented. Laser light allows for the transmission of very high energy pulses of light, at very high frequencies, and is crucial for modern high speed communication systems. This Nobel acknowledges critical work that was also simultaneously discovery (the maser effect) and invention (the maser and the laser), both central to the rise of the information and communication age. In Figure 1, the 1964 Nobel is also situated at the intersection of Invention and Discovery. The work on lasers built directly upon previous work by Albert Einstein but practical and operational masers and lasers where enabled by advancements in electronic amplifiers made possible by the solid state electronics revolution which began with the invention of the transistor.

Although scientists and engineers conducted a great deal of foundational work on the science of information technology in the 1960s, the next wave of Nobel recognition for this research did not come until the 1980s. Advancements in the semiconductor industry led to the development of new kinds of devices such as the Metal Oxide Silicon Field Effect Transistor (MOSFET). The two-dimensional nature of the conducting layer of the MOSFET provided a convenient avenue to study electrical conduction in reduced dimensions. While studying the _Hall Effect_ (an effect that describes the electric voltage produced across a current carrying wire when it is subjected to a perpendicular magnetic field) in the special case of semiconductor two-dimensional systems, von Klitzing discovered that the empirical measurements he was getting were exact in a way that had only been predicted approximately. Von Klitzing had discovered that given two-dimensional systems such as those found in highly refined transistors, in the presence of strong
magnetic fields and at very low temperatures, the Hall Effect is quantized. Instead of regular changes, von Klitzing observed steps and plateaus. The discovery was extremely useful in establishing electrical standards, as the values observed were so precise and accurate and in addition, could be expressed as integer multiples of two fundamental constants that they were used to define resistance. The quantized Hall effect can thus be seen as an important discovery with useful applications and is situated in the discovery half of Figure 1. It was also very important in catalyzing research into two-dimensional systems. For his work on the quantized Hall effect, von Klitzing was awarded the 1985 Nobel Prize.

The 2000 Nobel Prize was awarded jointly to Zhores Alferov and Herbert Kroemer and Jack Kilby. Kroemer and Alferov got one-half of the Nobel for "developing semiconductor heterostructures" and Jack Kilby got the other half for "his part in the invention of the integrated circuit." The 2000 Nobels can be classified primarily as inventions. The work on heterostructures built on the previous work done by Shockley et al.; again, we label these groups of Nobels a "family" because of their tight relationship and interdependence. This research enabled a new class of semi-conductor device that could be used in high speed circuits and opto-electronics. Early transistors were pretty slow and not fast enough to be used in high frequency circuits. Alferov and Kroemer showed that creating a double junction with a thin layer of semiconductors would allow for much higher concentrations of holes and electrons enabling faster switching speeds, and allowing for laser operation at practical temperatures. Their invention produced tangible improvements in lasers and light emitting diodes. It was the work on heterostructures that enabled the modern room temperature lasers that are utilized in fiber optic communication systems. We should note here that though Alferov and Kroemer’s work on heterostructures was recognized with the 2000 Nobel Prize, the actual work was carried out in the 1950s and 1960s and the consequences of this invention eventually led to the discovery of a new form of matter as discussed below.

Jack Kilby’s work on integrated circuits at Texas Instruments earned him his half of the Nobel for showing that entire circuits could be realized with semiconductor substrates. Shockley, Bardeen, and Brattain had invented semiconductor based transistors, but these were discrete components, used in circuits with components made for other materials. The genius of Kilby’s work was in realizing that semiconductors could be arranged in such a way that the entire circuit, not just the transistor could be realized on a chip. This invention of a process of building entire circuits out of semiconductors allowed for rapid economies of scale, bringing down the cost of circuits, and further research into process technologies, allowed escalating progress on the shrinking of these circuits so that in a few short years, chips containing billions of transistors were possible. The two inventions commemorated with the physics Nobel Prize of 2000 can be traced directly to the work carried out by Shockley et al. In Figure 1, the research honored by the Nobel Prize of 2000 is thus situated firmly in the Invention category.
Influenced by Alferov and Kroemer’s prior work, and with advancements in crystal growth techniques, (Molecular Beam Epitaxy which allowed for growth of layered heterostructures with atomic precision) Stormer and collaborators invented the concept of modulation doping where the charge carriers were physically separated from their parent donor atoms. This allowed for the fabrication of two-dimensional electron layers with mobility orders of magnitude greater than in Silicon MOSFETs. Stormer and Tsui then undertook the study the two-dimensional electrical conduction properties of these structures which showed unusual behavior. They cooled heterojunction transistors to a few degrees above absolute zero and exposed it to very strong magnetic fields. They discovered a new kind of particle that appeared to have only one-third the charge of the previously thought indivisible electron. Robert Laughlin showed through calculations, that this observation was a new form of quantum liquid where interactions between billions of electrons in the quantum liquid led to swirls in the liquid behaving like particles with a fractional electron charge. This beautiful phenomenon is clearly a new discovery, influenced by previous inventions with important practical applications (for example, high frequency transistors use in cell phones). For their work, Robert Laughlin, Horst Stormer, and Daniel Tsui were awarded the 1998 Nobel Prize in physics. In Figure 1, it is situated firmly in the Discovery category.

The next Nobel Prize in physics was in 2009, one-half was awarded to Charles Kao for “groundbreaking achievements concerning the transmission of light in fibers for optical communication” and the other half jointly to Willard Boyle and George Smith “for the invention of the imaging semiconductor circuit – the CCD.” Both prizes were directly influenced by previous inventions and discoveries in this area. Kao was primarily concerned in building a workable waveguide for light for use in communications systems. His inquiries led to astonishing process improvements in glass production as he predicted that glass fibers of certain purity would allow for long distance laser light communication. Of course, the work on heterostructures that allowed for room temperature lasers was critical to assembling the technologies of fiber communication. Kao, however, not only created new processes for measuring the purity of glass, but also actively encouraged various manufacturers to improve their processes in this respect. Working directly in industry, Kao’s work built upon the work by Alferov and Kroemer, enabling the physical infrastructure of the information age. Boyle and Smith continued the tradition of Bell labs inquiry. Adding a brilliant twist to the work Shockley et al. had done on the transistor, they designed and invented the charged coupled device (CCD), a semiconductor circuit that enabled digital imagery and later on, video. Kao’s work was clearly aimed at discovering the ideal conditions for the propagation of light in fibers of glass, but he also went further in shepherding the development and invention of the new fiber optic devices.
Figure 1: The Innovation Cycle in Information and Communication Technologies

(Note that the dates of the actual events are denoted in red)
From the discovery of the transistor effect that relied on the invention of the bi-polar junction transistor and led to all the marvelous processors and chips in everything from computers to cars; to the invention of the integrated circuit that made the power of modern computers possible while shrinking their cost and increasing accessibility; these six Nobel Prizes highlight the multiple kinds of knowledge that play into the innovations that have enabled the current information and communication age. The invention of fiber optics built on previous work on heterostructures and made the physical infrastructure as well as the speed of the global communications networks possible. In fact, the desire to improve the electrical conductivity of heterostructures led to the unexpected discovery of fractional quantization in two-dimensional systems and a new form of quantum fluid. Each of these could probably be classified as “basic” or “applied” research, but that classification elides the complexity and multiple nature of the research described above and does not help remove the prejudices of many against what is now labeled as “applied research”. Thinking in terms of invention and discovery through time helps reconstruct the many pathways that research travels along in the creation of radical innovations.

In our model, the Discovery-Invention Cycle can be traversed in both directions, and research knowledge is seen as an integrated whole that mutates over time (as it traverses the cycle). The bi-directionality of the cycle indicates the notion that inventions are not always the product of discovery, but can also be the product of other inventions. Simultaneously, important discoveries can arise from new inventions. The time dimension is captured in the idea of travelling the cycle.
CONCLUSIONS

In summary,

- All knowledge should be valued. Some production of knowledge is oriented towards improving our understanding of the world through the process of discovery; some is focused on the creation of new useful techniques and devices through the process of invention. The notion of the discovery-invention cycle is our attempt to view these two aspects of knowledge production as parts of a greater whole. Also by introducing new language, we hope to escape the cognitive trap of thinking about research solely in terms of initial motivations i.e., “basic” and “applied” but rather, consider the research process as contextual and generative along particular lines of inquiry.

- Technological visions and ideas that focus on the long term are of crucial importance, and can lead to both novel science and novel technologies. As demonstrated in our review of the Nobel Prizes for physics, and as further strengthened by the personal knowledge and experience of the lead author in the semiconductor field, we believe the issue of research time horizons — long versus short — to be much more central to explaining radical innovation than theories about the applications and motivations of scientific research.

- It is to be hoped that the discovery-invention cycle could be useful in identifying problematic bottlenecks in research. As it is essential to nurture every phase of innovation, this paper could be read as an argument that we must ensure the adequate alignment of various institutions to foster all aspects of the innovation process (namely invention and discovery).

- Successful radical innovation arises from knowledge traveling the innovation cycle. Bringing together the notions of research time horizons and bottlenecks, we argue that all parts of the innovation process must be adequately encouraged for the cycle to function effectively, the notion of traveling also emphasizes that we should have deep and sustained communication between scientists and engineers as well as between theorists and practitioners. Here, we hope to move beyond a singular focus on motivation, to argue that we must strive to bring all forms of research into deeper congress.

We have thus argued for an expanded alternative to the dichotomy of basic and applied research, for a paradigm shift to bridge the divisions between science and technology, and for a more holistic integration of knowledge as it travels the discovery-invention cycle. Our studies suggest that such a re-
thinking is necessary for the design of more effective research institutions -- where long time frames, a premium on futuristic ideas, and feedback between different elements of the research ecosystem are essential ingredients. This is especially pertinent in the case of the mission-oriented agencies such as the Department of Energy, the National Institutes of Health, and others.

**Implications for Research Policy**

The pertinent question therefore, is how do these insights play out in the messy world of actual policy making? First, there an obvious need to go beyond the unhelpful distinction between basic and applied research and toward a more integrated view that considers research holistically. The discovery and invention cycle draws attention to the entirety of research practice and allows one to pose the question of public utility to an entire range of activities.

Second, there is a need to explore more completely the nature of the public good, and thus the appropriate starting point for finding a role for the federal government changes. The binary of basic and applied was useful in one way, it provided a clear litmus test for limits to federal involvement in the research process. The idea that government funding of basic research leads to economic growth and is justified in “blue sky” projects that aren’t easily able to attract private funding is a useful one, but how does it translate when considering the discovery and invention cycle? We suggest that the federal funding is most appropriate for research that focuses on long term projects with clear public utility. The difference is that such research should properly lead to both new knowledge and new technology. The public good is best served by attending to the full innovation research cycle with an eye toward long-term projects, expecting the rewards as both discoveries and new inventions.
REFERENCES


