Patents and Patent Term Restoration: Incentives for Pharmaceuticals

Stephen A. Kuchta
University of Connecticut - Storrs, steve.kuchta@uconn.edu

Follow this and additional works at: https://opencommons.uconn.edu/dissertations

Recommended Citation
Patents and Patent Term Restoration: Incentives for Pharmaceuticals

Stephen A. Kuchta, Ph.D.
University of Connecticut, 2013

ABSTRACT

Patent protection has proven particularly important to the pharmaceutical sector among R&D-intensive industries. Despite a common patent term for all inventions, the effective patent life for a new pharmaceutical is shortened by the regulatory approval process. The Hatch-Waxman Act of 1984 created patent term restoration to address the sector-specific erosion of patent incentives. The Act restores a percentage of patent time lost during the clinical trials process, extending the patent’s expiration date and delaying the entry of generic competitors. Restoration therefore introduces additional strategic considerations for the firm, as their investment decisions affect the time spent completing clinical trials. This dissertation investigates these private incentives and the welfare tradeoffs created by incentivizing pharmaceutical R&D investment through patent term restoration.
Patents and Patent Term Restoration: Incentives for Pharmaceuticals

Stephen A. Kuchta

Ph.D. Law, Tilburg University, Tilburg, The Netherlands, 2010
M.A. Economics, University of Connecticut, Storrs CT, USA, 2004
B.A. Physics, State University of New York at Geneseo, Geneseo NY, USA, 2002

A Dissertation
Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy
at the
University of Connecticut

2013
Patents and Patent Term Restoration: Incentives for Pharmaceuticals

Presented by


Major Advisor

Thomas J. Miceli

Associate Advisor

Richard Langlois

Associate Advisor

Kathleen Segerson

University of Connecticut

2013
ACKNOWLEDGMENTS

This research is the product of an impossible number of people, a reckless amount of travel, and a combination of disparate academic perspectives. Every dissertation is unique but I hope that my particular route through the process lends as much beneficial character to the research as it did to me personally. To everyone involved, your thoughts, criticisms, musings, and jokes made this possible. Thank You, Danke Schön, and Hartstikke Bedankt. Specifically, I want to thank:

*My Parents,* whose patience I certainly tried but whose love and support remained unconditional. I owe everything to you both.

*My Colleagues,* To Nick, for never failing to have the answer and the patience to explain it; Marius & Inga, for keeping my head above water when I needed it, personally and professionally, from Vilnius to Willimantic, Frankfurt to Long Lake, & Switzerland to London; Rim, for introducing me to a new continent; Phil, for topping up everyone’s spirits; Rosanne, for always being there for all of us Grads in a thousand capacities; Kasey, for the hundreds of favors granted and gummy bears donated; and Craig & Josh, for helping an exhausted roommate have more fun than any student should.

*To Jonathan, and Rianne in Tilburg,* for time, support and understanding when I bit off too much at once;

*To my many teachers:* Ms. Sucher, Dr. Biaeve, Mr. Olson, and Mrs. Fuller; Pogo, Dr. Annala, and Dr. Stone; and Renate Seitz at the Baden-Württemberg Exchange Program. And of course, thanks to my Professors and my Committee here at UConn.

*To those I got to share the journey with:* Erik, Rob & Suus, Nicole, Selini, Pieter, Malini, Lorena, Bob, Robbie; Nate, Galati, Brady, Clif, Nicole, Wilson, Clader, Reingold, Gavin;
Bill, & Katie, Chen, Max & Shaina..., you're all in here too. Thank you for everything.

Finally, to Kat, who saw me at my worst but kept seeing through it to my best. You are perfect.
To Clarity
## Contents

Ch. 1. The History of Patents ................................. 1
  1.1 Philosophy and Narrative .............................. 1
     1.1.1 Enlightenment Evolution .......................... 6
     1.1.2 Adoption in America .............................. 9
  1.2 Modern Justifications ................................. 12
     1.2.1 Value Added ...................................... 16
     1.2.2 Focus on Pharma .................................. 19

Ch. 2. The Economics of Patents ............................ 21
  2.1 Hurdles and Claims .................................... 24
  2.2 Patents in Economic Theory ........................... 32
     2.2.1 Patents for Productivity ........................... 35
     2.2.2 Two Levers: Time and Scope ........................ 40
     2.2.3 Cumulative Innovation ............................ 46
  2.3 Licensing and Coordination ............................ 53

Ch. 3. The Pharmaceutical Industry ......................... 60
  3.1 Traditional Pharmaceuticals ........................... 62
     3.1.1 The Evolution of Big Pharma ...................... 68
  3.2 Biotechnology ........................................ 74
     3.2.1 Manufacturing .................................... 77
  3.3 Biologics and Biosimilars .............................. 82

Ch. 4. Regulatory Landscape ................................. 89
  4.1 Approval Pathways ..................................... 92
     4.1.1 The NDA and Clinical Trials ....................... 94
     4.1.2 The ANDA ......................................... 102
     4.1.3 The 505(b)(2) NDA ................................ 110
     4.1.4 The Paper NDA ................................... 111
  4.2 Biologics Regulation .................................. 112
Ch. 5. **History of Restoration** 120
  5.1 Historical Precedent & Future Pessimism .......................... 122
  5.2 Legislative History .............................................. 124
      5.2.1 The Restoration Calculation ............................. 132
  5.3 The Refund Rate ............................................... 136

Ch. 6. **PTR, Diligence, & Welfare** 141
  6.1 Two Stages of R&D ............................................. 142
      6.1.1 The Clinical Period ........................................ 144
      6.1.2 Research Incentives ...................................... 149
  6.2 Optimization and Reactions ................................... 151
      6.2.1 Social Welfare and Policy ............................... 153
  6.3 Discussion .................................................... 154

Bibliography 158
  References ......................................................... 158
Chapter 1

The History of Patents

One can hardly escape from the conclusion that a lawyer who has not studied economics and sociology is very apt to become a public enemy.

Justice Louis Brandeis

1.1 Philosophy and Narrative

Patents are incentives for efforts aimed toward tasks that are uncertain in outcome. Success manifests in the invention of new items or innovation of existing ideas, and as such has value to individuals and society.¹ Unfortunately, by the nature of knowledge, innovations can often be copied at lower costs than that incurred by the inventor. A patent therefore works to incentivize uncertain activity by promising security from copying after success. Thanks to the patent, society gains a product or idea that otherwise might not have the intrinsic

¹Fundamental treatments of patents, innovation, and economic growth begin with Arrow (1962).
financial justification to be developed.\(^2\) This view of the incentives offered by patenting can be seen as the core of an "ex ante" justification for the patent system (Lemley, 2004).

The perspective on patents changes however after successful invention and market introduction, when "ex post" justifications come in to support the narrative (Lemley, 2004). Patents cause the textbook social detriments associated with market power, a combination of increased prices and restrained output that transfers social surplus to the patent holder. These profits serve as the ex post reward for the uncertain ex ante investment. For the modern patent system, that detriment is limited in duration, after which the competitive market gains control of the new knowledge.\(^3\) The economic rationale for a patent system therefore rests with the prediction that absent such a reward, fewer inventors would spend their time inventing and those that do would have reason to keep the content of their inventions secret. The former situation reduces the prospects for, and rates of, economic growth while the latter brings the same detriments of market power without the benefit of expiration.

Although the ex ante and ex post justifications have taken on more nuanced characteristics, a combination of these basic tenants runs through the patent narrative since the dawn of intellectual property in 14\(^{th}\) century England (Mossoff, 2001, p. 1259) and 15\(^{th}\) century Venice (Schaafsma, 1997; Mandich, 1948). In Venice, skilled craftsmen were granted monopolies over the practice of their particular technique so long as they taught the craft to others (Schippel, 2001) or provided a service to the state via their craft (Mandich, 1948). Early awards as far back as 1272 find the Cabinet of Venice granting to foreign woolweavers a house to live in and a shop to sell from, free of charge, for 10 years (Mandich, 1948, n. 17, p. 171). A predictable stream of profits then went to the uniquely skilled crafts-

\(^2\)This work takes the patent grant as a primitive concept. More general questions of innovation policy are explored by Penrose (p. 27, 1951), among many others, who entertain whether monopoly grants are the correct way to fulfill any implied duty to inventors.

\(^3\)To 20 years among WTO members who have signed the General Agreement on Trade and Tariffs. For an introduction to the global protection of intellectual property from the pharmaceutical perspective relevant for this research, see Kyle & McGahan (2012).
man who controlled the market in that trade.⁴ Although with hindsight one might argue that difficulties in replication in craft industries may well have lent market power even where technique could be reverse-engineered, the success of the industries with monopoly protection and the relative growth of cities and states with more developed property rights systems lent momentum to the practice. The problem of causality, however, persists and was captured in Edith Penrose’s (p. 40, 1951) prescient statement that

“[i]f national patent laws did not exist, it would be difficult to make a conclusive case for introducing them; but the fact that they do exist shifts the burden of proof and it is equally difficult to make a really conclusive case for abolishing them”

The static burden of patent systems is the market distortion born by society, under the assumption that an influx of new practitioners would drive down prices. What incentive though would remain for an innovative tradesperson to reveal their new techniques? One might argue for a resort to trade secrets, protecting the inherent knowledge of any product while still reaping market profits. Even where possible, overall economic growth suffers where knowledge is hidden. Therein lies the dynamic problem of innovation policy: where the nature of the innovation would allow profiting through secrecy, society would still be better off with the new information and the new invention than with the innovation alone.

Jeremy Bentham (Bentham, 1843, Chap. 3, p. 71) summarized the situation in stating “[h]e who has no hope that he shall reap, will not take the trouble to sow […] [b]ut that which man has invented, all the world can imitate.” Thomas Jefferson, acknowledged as the father of the American patent system, echoed “[h]e who receives an idea from me, receives instruction himself without lessening mine; as he who lights his taper at mine, receives light without darkening me.”⁵ Such has been the justifying narrative for patents up to

⁴A 10 year term in the case of new and ingenious inventions, although it could be longer, in excess of 50 and 25 years, depending on the situation at hand (Mandich, 1948). One is reminded of the modern practice of granting firms tax breaks and waivers for new construction or locating in a certain state or town.

⁵Graham v. John Deere Co., 383 U.S. 1, 9, n. 2 (1966)
the modern day and persists despite modern doubts and empirical research questioning the stagnation of progress in a world absent protection.6

Causality is a fickle inquiry, but patent protection is, at a minimum, correlated with stronger economies (Lerner & Merges, 1998). The correlation was well-used by the guild system in Renaissance Italy, which lends the location’s usual title as the beginning of the modern patent system. Guilds negotiated with the Italian city-states and amongst members, culminating in formal rules for settling disputes and governing members’ behavior in the marketplace (Schaafsma, 1997). The guilds recognized the danger of displacement—creative destruction—manifested by new inventions. Unfortunately, Venice focused the destructive energies on the innovator in allowing for “destruction of infringing devices and payment of a fee to the inventor (Schaafsma, 1997, p. 244). Venetian patents were therefore not merely an ability to exclude competitors but a de facto license to operate where the government was both the granter of patents and the active arbiter of market competition (Mandich, 1948).

Despite literal destruction of infringing manufacturing, the State does not—neither then nor now—formally exclude individuals from accessing the information within the patent itself.7 In the historical examples, information transmission came through the condition on teaching the patented process to apprentices (Grubb, 2005, p. 6). In fact, the word patent itself derives from the teaching dynamic (Mossoff, 2001, p. 1259). The original term of letters patent—literae patentes— or open letters are so called “because they are not sealed up, but exposed to open view... (Blackstone, 1768, p. 316-317).”8 Protected skills and industries were therefore a compromise, a balance, where society received both a new invention or

---

6For recent discussions see Hope (2008), Bessen & Meurer (2008) and Boldrin and Levine (2008). For comments on other types of motivation, especially incentives from bazaar-style production, driven by use value, see Hope (2008). The term “bazaar” refers to nonhierarchical organization and references Eric Raymond’s The Cathedral and the Bazaar, (2001). More traditional economic questions of strong patents may be found in, i.a.; Mazzoleni & Nelson (1998).

7De facto exclusion via the inability to experiment with an infringing device notwithstanding.

8A sealed up patent is a trade secret, which actively controls the information (D. D. Friedman, Landes, & Posner, 1991).
product, received instruction in its use both formally through instruction and informally through the existence of the product itself, and the inventor received an inter-temporal incentive, allowing them to profit from inventive activities.

This narrative continues today although much has changed about innovation and industry. For instance, the industrial structure at the beginning of patents was restricted to craft techniques, and hence the reliance on guilds. There was no mass or mechanical production that would enable truly inexpensive replication of the fruits of another’s labor. Hand production and skilled craft work is not unambiguously capable of eroding the profits of the innovator enough to deter innovation (Mandich, 1948, p. 166). There remained however the threat that copying might result in displacement, as a follow-on innovator entered the market with what is known today as a non-infringing competitor. Nevertheless, if there were no incentives to get the ball rolling, the pace of invention might suffer.

The goal of beginning a discussion on patent term restoration with the historical context of general industrial patents is to highlight the continuity of what one could call the patent narrative. Patents were argued into existence as a reward for successful innovation. The reward finds justification as an inter-temporal incentive correction, bringing profits back to the inventor from their future market. The State, for their part, sees growth from the inventive activity and finds some way to balance the private incentivization via the social benefits. At the start of modern patenting, the balance turned on teaching the invention to the next generation of practitioners. The modern view balances private profits with information disclosure and some more nuanced legal frameworks which allow for other inventors to compete with inventor as long as their product does not infringe the intellectual content of the innovator’s product. That is, the patent does not explicitly protect the market now but rather the intellectual contribution.

In sum, we find patents first supplying a reward or remuneration to the innovator for their contribution. Later, the reward expires and leaves a new market landscape with additional products and new information for consumers and innovators. While a coher-
ent narrative, patents have yet to experience a modern consensus incorporating the tools of modern economics and finance. The patent narrative has remained remarkably static through both the political upheaval preceding the modern era and the industrialization of the globe which followed.

1.1.1 Enlightenment Evolution

The patent narrative begins at the connection between business interests and State involvement, introduced above with reference to Venetian guilds, and departs from there in the direction of economic growth. Even if one assumes, for the sake of argument, that the positive outcomes of incentivization will at some point outweigh the temporal detriment of awarding market power, the narrative quickly runs into another existential question. Government has had a rather long and unsavory history of promoting friends at the expense of others. Tacitly agreeing to provide innovation incentives should immediately engage the legal scholar in an effort to restrain bureaucratic moral hazard. Success was not achieved until the modern age, cordonning the economic benefits of the patent behind the portcullis of objectively-promulgated intellectual property regulation. As with all areas of law, however, objectively-stated criteria leave much room in practice, resulting in much argumentation as to who and what gets through the gate, and questions as to what members of the populace already residing inside the gates have in common with new entrants.

As law is an evolutionary phenomenon, we may state broadly that the benefits that flows of new knowledge have lent to a governed region should be expected to favor looser entrance requirements. The benefits helped a young Britain so that they issued patents to those who went out into the world and brought inventions back to the homeland. The ideas they need not be invent by the importers themselves (Grubb, 2005). Patents could also issue for re-establishing industries within the realm. Where a trade was no longer practiced, the

---

re-establishment would be economic growth, a justification built on the Renaissance notion that patents bring foreign craftsmen (Mossoff, 2007, p. 1259). The 16th century English letters patent notably came with the requirement to set up shop in England and carried a positive duty on the recipient to practice the trade (Mossoff, 2009).

Though a loose entrance requirement, the realizations tended to favor interests of those in power. Penrose (p. 17, 1951) notes that “the inventors’ privilege was not given indiscriminately as a matter of right, but selectively to encourage or make possible the development of specific products or processes which were considered of economic importance to the state.” True to this philosophy, while importing industries was worthy of Crown protection, manufacturing patents could not issue when they would compete with existing industries, as that would just be displacement and not growth (Mossoff, 2001). Naturally, the sovereign did not always follow economic logic.

The story of Sir Robert Mansell is illustrative. In the early 17th century, Sir Mansell held a patent on glass making using coal as a heat source. The characteristics of the fire in producing glass required a slightly different method than using timber. King James I, ostensibly to conserve timber as a resource, dictated that coal must then be used in the glass making industry. The result was a predictable transfer of wealth from the extant glass industry to the patent holder (Brimblecombe, 1987, p. 33). Sir Mansell’s political power precipitated the change and not any—if there was any—personal contribution to the invention, let alone any rigorous economic analysis of the change in input technology (Bond & Price, 2006, p. 81).

Sir Robert Mansell’s story, illustrates an awareness of the reward aspect of a patent. Commentators have summarized that patents at the time had evolved into a form that “had nothing to do with legal rights or even inventions per se, but rather they represented royal privileges that supported royal policies (Mossoff, 2001, p. 1261). This is not entirely

---

10England at the time saw this as a counterweight to the medieval “brain drain” towards the continent (Mossoff, 2007). Also see Ibsen (2012).

11Recounted in full by Carroll and Price (p. 67-81,217-241, 2006)
unexpected; at that time, “the organization of municipal and business life was largely based on special charters, privileges, franchises, and licenses, and the patent of innovation was not easily distinguished from other privileges granted by the Crown (Penrose, p. 4, 1951).” There is at least a component of “business as usual” on top of any formal corruption, although there were instances thereof (Penrose, 1951, p. 5-6).

While the observer—cynical or not—readily sees the shadowy influence of power, there must have been sufficient areas of gray, where benefits were at least on balance with corruption, because monopoly abuses did not end with the decline of monarchy power in Britain.  

In fact, they did not end with the rise of the independent industry. At a time where unearned privilege was under intense scrutiny, the patent as an institution survived. Short of entertaining a larger conspiracy, or diverting too far down a separate history project, we may assume that the problems of the patent system were on average seen rectifiable, and perhaps easily rectifiable, within the legal paradigms emerging under the Enlightenment project.

The emergence of a property rights-centered justification for patents, arriving with the Enlightenment milieu, did not necessarily translate into different behaviors on the part of innovators and government however. A form of mercantilistic appropriation remained, nurtured by the emergent industrial business class. The tangible benefits of growth-through-innovation can be seen as, at a minimum, providing arguments for the patent system in an otherwise hostile environment. As Schaafsma (p. 243, 1997) notes,

[u]nlke most Mercantilist thought which has been discredited by modern economics... use of limited grants of exclusivity to encourage economic development

---

12The pattern of privilege granting in Britain led to a conflict in 1571 between Parliament and the Crown (Mossoff, 2001, p. 1262) and again in 1624, leading to the passage of the Statute of Monopolies, which removed most of the the ability of the crown to issue grants (Schaafsma, 1997).

13Boldrin and Levin (Introduction, 2008) relate the story of patent blocking and rent seeking behind James Watt’s steam engine patent, which began in 1775 and extended by Parliament through 1800.

14Schaafsma (p. 243, 1997) explains “[i]t was not surprising that the use of governmental grants of exclusivity to encourage economic development arose at a time when merchants were putting forth various economic principles designed to support their self interest.”
remains a viable economic tool.

Moreover, the patent system which remained in Britain through political changes\textsuperscript{15} also successfully booked passage to the New World.

1.1.2 Adoption in America

Patents found fertile ground as the young U.S. Congress built their patent practice on old precedents. The U.S. Constitution authorizes patents based explicitly on the philosophy of promoting intellectual progress.\textsuperscript{16} Under such high-placed authority, the United States Patent Act of 1790 created the rights to exclude others from newly-created intellectual property, setting patents as an exclusionary power and separating it from sibling forms of intellectual property, such as copyrights. It does represent a firm break from the historical positive requirement to practice the art, which is perhaps the most substantial modification of the practice and the intellectual arguments surrounding patents to emerge formally from the progressive era.\textsuperscript{17}

The rights of the individual also received renewed focus, and the U.S. patent grant starts with a local inventive hand; it cannot not begin with importation. The stamp of the labor theory of property and value is evident in the movement away from mercantilism (Mossoff, 2007, p. 982).\textsuperscript{18} Though intellectual changes redefined some edges of the patent

\textsuperscript{15}See, i.a., Grubb (p. 8-10, 2005) and Mandich (1948).

\textsuperscript{16}Article I, §8 of the United States Constitution: "The Congress shall have Power... To promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries." Lemley (1994) notes that the conflict in independent state-based patent systems in existence prior to U.S. independence drew the power to grant patents toward the federal government. See Mossoff (2007) for a full discussion on the evolution of patents as rewards or privilege with respect to the U.S. Constitution. See Ochosa (2002) for a history of similarities between patent and copyright extension legislation in the young Congress, relative to both the current discussion and the coming discussion of patent terms.

\textsuperscript{17}Copyrights are an affirmative right for the owner to do or authorize activities with. See (Hope, 2008, p. 165). As a separate topic with separate legal underpinnings they are largely absent from this discussion.

\textsuperscript{18}Edith Fenrose (p. 21-26, 1951) also discusses the prevalence of natural rights philosophy, such as the debate over the International Convention for the Protection of Industrial Property. It is also evident in the early quote from President James Madison: "[j]ust as a man may be said to have a right to his property, so he has a property in his rights." The Papers of James Madison, Robert A. Rutland, Thomas A. Mason, Robert J Brugger, Jeanne K. Sisson, and Fredrika J. Teute (eds.). Charlottesville: University Press of Virginia,
system, the discussion did not penetrate into the workings of the patent, as evidenced by the static length of the patent term. The prevailing historical statutory length of 14 years, which was said to arise from the length of time required to teach two generations of apprentices (Grubb, 2005, p. 6) survived the Atlantic journey in tact (Walterscheid, 2004, p. 600,603-04).19 Walterscheid (p. 600, 2004) relates

“[t]he initial statutory terms of the patent and copyright grants were simply copied from the existing British law, both because this was the easiest thing to do, and because Congress had no basis on which to make any determination that American conditions were sufficiently different from British conditions to justify the setting of different lengths of the term. Thus the [U.S.] Patent Act of 179020 authorized the issuance of a patent ‘for any term not exceeding fourteen years’.”

Although one might argue such policy inquiry would have been a stretch for the young Congress of the time, Grubb (p. 13, 2005) relates how that same Congress discussed developing the arts and sciences instead by granting cash awards. Such practice would be a true award rather than the tangle of reward, incentive, and desire for economic growth contained in patents. Cash awards, however, would have obviously cost money, while granting a patent does not. The young Congress may have been willing to wade into the intellectual discussion of optimal or efficient innovation policy but they did not have the luxury of funding such policy.

After enactment, and despite objections of those who saw patents as the same “odious monopolies” left behind in England (Mossoff, 2007, p. 959,n. 32), the patent developed through momentum borrowed from the timely arrival of the Industrial Revolution in Americ-

19One can add to that the anecdote of Thomas Jefferson, as father of the USPTO, and his slave and cook, James Hemming. James was schooled in French cooking when Ambassador Jefferson brought him to Paris. Upon returning to the U.S., James asked to be freed, a condition Jefferson agreed to once Hemming passed to another slave the skills he learned in France.
ica. By the end of the energetic revolution, patents had cemented their present theme as rewards to the worthwhile inventiveness of the citizenry (Grubb, 2005, p. 13). They had certainly supplied a new world full machines as anecdotal evidence of effectiveness. At the very least, a mountain of progress during a time when the patent system was readily employed by business would have to be addressed by any research hypothesizing a net downside to patent protection. In fact, it took until the middle of the 20th century for economists to seriously consider the backstory of patents and empirically question their role.21

Economic interest in the subject arrived on the same concerns though that marked the early incarnations of the patent narrative. While a more accessible legal system modernized the patent system and protocols for their award, modern industrial interests are often seen exploiting even the tighter protocols. They use patents and the Courts to define, redefine, and enhance their areas of influence, playing games with law that focuses on subjective intellectual components and does little to recognize—or utilize—the economic outcomes which may cast reveal objective boundaries for the subjective arguments. Although the economic inquiry revealed much, there has been no major change to the patent incentive. As Engelberg (p. 394, 1999) summaries that “there was no legal or logical relationship between the life of a patent and the commercial life of any product claimed in a patent” and that statement remains true.

The incentives for innovating under the modern patent system are formally introduced below. The historical narrative colors the modern story and generally finds the patent narrative insufficient to justify the modern patent system (Penrose, 1951). Modifications of the one-size-fits-all patent, built on economic rationale over legal precedent, remain a significant hope for an optimized invention incentives system. The existing modifications of patents in the pharmaceutical industry therefore serve as an outlook over the general

---

21Fritz Machlup, in the preface to Edith Penrose’s The Economics of the International Patent System (1951) bids “Welcome Back” to economists who, in his understanding, have been absent from the patent debate since the debates of the 19th century ended with the (Paris) International Convention for the Protection of Industrial Property of 1883.
landscape.

1.2 Modern Justifications

With the influx of economic inquiry, the patent narrative developed more formal language. Most terminology arrives from existing economic fields. Before turning to the literature proper, we use the translation of the historical narrative into modern parlance as an illustrative introduction.

Bridging ideas start with public goods. Arrow (1962) described knowledge as the quintessential public good and the non-rival characteristics were further emphasized by Barzel (1968). Classically, public goods cause problems for markets because they are non-excludability and non-rival in nature. With textbook examples—lighthouses and national defense—a central authority utilizes taxation to fund the beneficial expenditures where the market would under-supply. Free-riding is then foreclosed by the State's power to tax. To make the analogy more relevant, we may point to the grants and scholarships routinely made by modern governments through the funding of universities and related research endeavors.

Where the public good analogy breaks down is in the uncertainty. The construction of a lighthouse, and the social benefits which follow, has a certain outcome; research does not often deliver a well-defined outcome. Appropriating funding for research, broadly defined, therefore requires additional justification. There are of course ways to define a research project so that it does have a well-defined goal and therefore delivers an outcome on which the taxation was justified and, presumable, to which the taxation was agreed. The uncertainty would then be incorporated into the decision from the start. Difficulties exist however with a central authority's expertise in any innovative area. Hayekian (1945) arguments might be over-powered here, but the diffusion of knowledge in a society is unlikely to be grasped and controlled, even in the limited sense of a single research project, by the
central authority.\textsuperscript{22} The diffusion of information in society and the amount of un-codified knowledge leaves even those with the expertise in a difficult position deciding on any forward focused research plan.\textsuperscript{23}

The most important point of the public good comparison is the recognition of the agglomerative role a central government may perform. The role is refined here by placing government in the position to exclude individuals not from from accessing, for example, the lighthouse, but by preventing them from profiting from using the lighthouse. With intellectual property, information is the light from the lighthouse and although a ship may benefit from the construction it will be prevented from docking and accessing markets for its safely transported goods. And to stretch the analogy, the blockade has benefits for the excluded ships too.

By the nature of knowledge, it grows in value even when not engaged directly in the market. Only the holder of the patent may be profiting from a given invention but the rest of society is learning from its mere existence. Thomas Jefferson refined this statement, stating

\begin{quote}
\"[u]nlike the case of lighthouses, the government has a rather effective way to exclude potential free riders. After all, the information is not exclusively what the free rider is after. Rather the transformation of that costless information into a revenue-generating marketable product is the desire.\"
\end{quote}

On the other side of public goods and the shipping metaphor, economists recognize the potential for a \textit{tragedy of the commons}, or the overuse of shared good, held in common.\textsuperscript{24} Property rights can be the solution to the overuse problem, adding ownership and a private cost to marginal usage. While knowledge is not rival, patents can theoretically over-divide

\begin{flushleft}
\textsuperscript{22}For an introduction to centralized incentives, like prizes, as innovation policy see Wright (1983).
\textsuperscript{23}For discussion on the economics of research prizes, see i.a., Wright (1983), Denicol\`{o} (2000) and more recently Abramowicz (2003)
\textsuperscript{24}See Hardin (1968) for the first mention of the tragedy, leveraging an environmental example. For pharmaceutical discussions see Heller and Eisenberg (1998), Adelman (2005); and Mireles (2005).
\end{flushleft}
property interests in knowledge, resulting in an anti-commons effect (Heller & Eisenberg, 1998).

A balance is needed in innovation policy, a balance which distills many positives into the general phenomenon of incentivizing surplus-generating innovation, and an equally numerous spectrum of social detriments created by offering time-limited market power awards as that incentive. For a single or solitary invention, defined as one which does not draw inspiration from other inventions and moreover where we proscribe the conversation from any inspiration effects on subsequent inventions, the simplest balancing that of adequate remuneration for up-front expenditures. The patent secures an expectation of economic profits after invention and policy balances the incentive by limiting the profit period to just incentivize the R&D expenditure. That is, the patent term tuned so that the market profile of the invention generates sufficient profits to compensate the inventor for their ex ante effort and costs. Society benefits from the invention as soon as it enters although to a lesser extent than it would if competition would allow. Allowing that competition into the market once the innovator has been compensated completes the cycle and the goal is accomplished at a minimum—though non-zero—social cost.

The assumptions are quite stark however. The limited legal evolution away from the mercantilist tradition puts the modern patent in a position of protecting only the content of the innovation. There is no guarantee of a monopoly position in the newly-created market and where one does exist there is no guarantee from the patent that the power position will last for the duration of the patent. The knowledge dissemination may block one direct entry into the market, but an indirect approach is not foreclosed. A second inventor may study a patent and develop an entirely unrelated innovation, bringing a second set of information to bear on a problem. Whether or not the original information catapulted the second inventor into a position to compete is irrelevant to the legal side of patenting; it is clearly important for the economic profits which provide the incentive to innovate in the first place.
A patented invention may contain a spark, catalyzing subsequent development and revealing directions previously hidden. The more information is disseminated, the more society benefits from the fruits of subsequent innovation. Conversely, the more cumulative activity, the more trouble breaking even an innovative inventor faces when opening a new market.\textsuperscript{25} The threat is known as “inventing around” an original patent, where followers step up, over, and around the original in a cumulative innovative processes (Scotchmer, 1991). Society sees gains as competition in the knowledge-space opens up more and more market space. The benefits of the expanded, cumulative, patent narrative depend on keeping the balance though, helping incentives up to the task of surmounting the new type of competition. The patent has effectively created new incentives while trying to balance the original tradeoffs.

The view from economics is then of an incentive policy, grounded on the social benefits and economic growth promised by innovation, which attempts to balance the private and the social with a minimum of fallout. The fact that there are benefits to tuning the patent to the market characteristics, including the eventual knowledge-competition which may be unleashed by incentivizing a solitary invention, is not lost on the economics literature. The economic perspective though runs into the one small directional change that the patent narrative has made in its historical evolution, where the patent was connected to the intellectual contribution and removed from market considerations. The spark-effect of patenting would seem to indicate that society can suffer more detriment and still come out ahead though. The next section adds more perspective on patent’s secondary effects, locating some of the rationale used to support expanded patent incentives in the face of cumulative, dynamic, innovative processes.

\textsuperscript{25}Compare however Scotchmer (2010), Maurer and Scotchmer (2006), and Lemley (2002a) on the limited amount of disclosure in software patents, an area with tremendous growth despite limited disclosure.
1.2.1 Value Added

The benefits of legally defining intellectual property fall generally into four categories, summarized by Nelson and Mazzoleni (1998). The first is this “inducement” rationale highlighted above where the expectation of private property grants motivates innovative activity. Here, longer patents are larger inducements, providing greater expectation of profits even in the face of uncertainty through simple time increases alone.

The second category of benefit lies in “disclosure” theory, also introduced above. Disclosure theory posits that the benefits of patenting arise through the release of information which would otherwise remain secret. The social benefit of disclosure is codified in modern patents with the protocol for applications, namely in the publication of patents soon after their award. Note however that the disclosure happens anyway in the case that the product comes to market. Therefore, the social consideration found in disclosure exists primarily in cases of inventions that never come to market, or would be difficult to understand without the help of the patent as an explanatory device (Grubb, 2005, p. 10). As a part of the application protocol however we can see that a policy requiring greater disclosure or specifically, greater teaching, might be exchanged for stronger patents.

The third benefits rationale builds on the inducement theory and adds that patents secure the investment necessary to pull inventive ideas into the commercialization stage. A patent as intellectual property makes it credible to a funding source that an inventor will maintain any advantages to their invention. The patent therefore incentivizes invention not

\[26\text{One should note that this is all viewed from the “modern era” of economics, beginning in the 17th century and the collapse of the feudal order of society. Prior to that, in the Scholastic Economic period, private property found justification by such thinkers as St. Thomas Aquinas in aspects of fairness, especially to seal the concept among biblical condemnation of private property ownership (Schaafsma, 1997, p. 242).}

\[27\text{The history of patents shows that the idea of disclosure became the “consideration” for the grant of the patent (Grubb, 2005, p. 10).}

\[28\text{The requirement of disclosure is enunciated in 35 U.S.C. §112, under the heading of specification; “The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same, and shall set forth the best mode contemplated by the inventor of carrying out his invention. The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.”}
only because of abstract ownership incentives to the inventor, but in the practical sense through the realities of R&D finance. The tradeoffs here are again straightforward, with increasing benefits to patents justifying more complete or more secure ownership, which in turn enables the formulation of more exact financial expectations.

The fourth rationale is Kitch’s (1977) prospect theory, which introduces a patents’ ability to efficiently order follow-on markets. Paralleling the ability to secure R&D funding, granting ownership to ideas bestows incentives to actively work the invention (Clarke & Kohler, 2005, p. 12). To the extent that first movers are well-placed to coordinate subsequent activity, due to latent, uncodified knowledge or specialties developed in the preliminary stages, then the patent lends momentum to the forward motion. Where inventive prowess stops at the idea stage, the ownership allows efficient transfer, allowing the innovator who has taken an idea to the edge of their knowledge to pass it on to someone who can develop the idea further. By placing ownership in the hands of those involved at the ground level of invention, society establishes incentives towards the alignment of development resources.

There are, of course, downsides. Prospect-theory’s allusions to a gold rush could also lead to analogues of the less-efficient outcomes highlighted by the namesake (Haddock, 1986). Premature capture can lead to purely redistributive outcomes (Clarke & Kohler, 2005, p. 112-13). In the patents literature, this is the idea of “Rembrandt’s in the Attic” (Rivette & Kline, 1999), where one stockpiles intellectual property in the hope that it will become valuable at some later date. There is no claim of efficient ordering of follow-on research here. Intellectual squatters can now block efficient development, extracting tolls.

[29] Modern pharmaceuticals often emerge from venture-funding projects, where the only asset the initial firm has is the intellectual property over a molecule. See i.a, Friedman (2008) and Gassmann (2008)
[30] Also see Epstein (1986)
[31] Note the entry bias found by Mankiw and Whinston (1986), where entry is more valuable to the entrant than to society when their entry induces lowered output of incumbents, in that the individual soaks up surplus.
[32] As Haddock (p. 778, 1986) explains, the tipping point is “the measurement problem—the definability of a resource before it is ready to be exploited. Some resources have insufficient value today to tempt anyone to bear the present costs of establishing and enforcing title, but are recognized to be of increasing value in a growing economy. […] Other resources cannot even be described at present.”
along a path that would otherwise have been open to the suitably-knowledgable inventor.\textsuperscript{33} Navigating this "patent thicket" is, by most measures, a non-trivial process (C. Shapiro, 2001).

These four paragraphs seem circle around toward a disheartening conclusion. While pursuing quantifiable mechanics for social benefits of patents, we tend to uncover more axes requiring balance. Anecdotes become particularly worrisome in the sciences, where it is easy to find problematic examples in the world of gene patents (Kane, 2004). If genetic patents creates a patent thick, it is a thicket which may absolutely prevent "inventing around."\textsuperscript{34} On the other hand, research also locates support for prospecting's ability to align research efforts, finding evidence of the innate industrial ability to coordinate (J. P. Walsh, Cho, & Cohen, 2005; J. P. Walsh, Arora, & Cohen, 2003; J. P. Walsh, Cohen, & Arora, 2003).\textsuperscript{35}

In addition to providing a timely anecdote, the eventual therapeutic market created by gene patents, directly or indirectly, is the only market in which the patent term is able to be modified.\textsuperscript{36} In all other sectors, the patent grant is entirely static.\textsuperscript{37} The adjustments

\textsuperscript{33}Though not developed in this work, firms engaged in R&I generally manage a portfolio of patents (Dasgupta & Maskin, 1987). While generally reflective of their area of expertise, a portfolio can be seen as either a help or a hinderance to coordination, providing both leverage to secure patents one does not own or complicating bargaining for others. Further complicating matters, the ownership of a patent is not always clear in practice. While the inventor is listed on the patent as the original owner—U.S. patents may only issue to people, not corporations—assignment of the patent rights may pass between subsequent ownership without notification of the issuing authority. The complexity of first identifying and then bargaining with the owner has the potential to generate the interesting if counterintuitive incentive to "ignore" patents (Lemley, 2008). Owing to a legal difference between willful patent infringement and accidental, it may prove efficient to, in effect, cover one's eyes and invent rather than continually check the patent publications for directions, a fact which naturally reduces the patent system's claimed benefits through disclosure.

\textsuperscript{34}Recent survey evidence from Langinier (p. 3, 2006) finding one fourth of surveyed laboratory physicians having abandoned development of a clinical test due to patent problems. Near nearly half of the respondents reported purposefully refraining from developing a test out of the fear of a lawsuit. Also see Louis, Jones, Anderson, Blumenthal and Campbell (2001). Studies going back as far as 1986 and extending to today have shown that data can be withheld between colleagues because of concerns that the sharing will impact future commercial prospects (Blumenthal, Gluck, & Seashore Louis, 1986; Blumenthal, 1992; Blumenthal, Campbell, Causino, & Louis, 1996a, 1996b; Blumenthal, Campbell, Anderson, Causino, & Seashore Louis, 1997; Campbell, Weissman, Causino, & Blumenthal, 2000; Campbell et al., 2002; J. P. Walsh & Hong, 2003). See note commentary in Hope (p. 8, n. 14, 2008).

\textsuperscript{35}Coordination mechanisms exist in intellectual property transfer as they do for physical property, including the possibility of shared ownership or shared access through patent pools and standards setting (Lerner & Tirole, 2004; Merges, 2001; C. Shapiro, 2001). This research limits discussion to single ownership and solitary inventions.

\textsuperscript{36}Chapter 4

\textsuperscript{37}Chapter 2
are known as Patent Term Restoration (PTR), while remaining blind to the market profile created by the therapeutic, they do address the social need to provide safety and efficacy testing prior to entering the market. PTR came into existence to counteract the loss of effective patent life due to the necessity of regulatory hurdles. In the US, the Food and Drug Administration (FDA) is the gatekeeper and oversees tradition drugs, new biologic medicines, and medical devices. The level of regulation for the R&D endeavors of these firms is unique.\textsuperscript{38} Just as unique, PTR is the only exception to the one-size-fits-all patent grant.

\subsection*{1.2.2 A Focus on Pharmaceuticals}

The regulatory dimension reintroduces a central party concerned with the market for the innovative products. In contrast to the historical context, there is a physical and legal separation between patents and market gatekeepers, where each utilizes a different sphere of information.\textsuperscript{39} However, in line with economic arguments, there is a chance to tune patent awards for invention characteristics.\textsuperscript{40}

While a full consideration of the overlap in pharmaceuticals requires a discussion of regulatory exclusivity, this research focuses first on the PTR modification.\textsuperscript{41} Though couched in the legal terms of fairness, the operation of PTR reaches strongly to patenting’s historical narrative in its focus on adequate reward, via profits as reimbursement, for the efforts of the innovators.\textsuperscript{42} By its formulation as a reward for a single approved pharmaceutical, PTR,

\begin{enumerate}
\item\textsuperscript{38}Pesticides also a regulatory, though interact with a much smaller market.
\item\textsuperscript{39}in re Brana, 51 F. 3d 1560 (Fed. Cir. 1995).
\item\textsuperscript{40}Patent lengths in the United States are governed by the Patent & Trademark Office (PTO) who does not consider the FDA’s regulatory involvement when setting their terms. The FDA for their part does not generally pay heed to the outstanding patent situation, although there are certain exceptions dealing with secondary entrants (Chapter 4). Secondary entrants are known as generic drugs and may, to some extent, circumvent the first entrants regulatory hurdles by applying for abbreviated approval. For a complete discussion, see Derzko (2005).
\item\textsuperscript{41}The overlap of regulatory exclusivity as a market incentive and patent life as an incentive is introduced in Chapter 3. For complete discussions, see Dudzinski and Kesselheim (2008); Dudzinski (2005); and Zweiful (2004).\
\item\textsuperscript{42}Such concepts of fairness to innovators lives on in proposed legislation such as the Patent Fairness Act of 1999. (HR 1598, 106th Cong. (1999)). See Patel (2001). Morgan (2010) explicitly notes the confusion in
in the form of the Hatch-Waxman Act of 1984 (HWA), stakes its boundaries around the solitary invention perspective, well short of the economic pros and cons of viewing invention as a cumulative process. As such, there are clearly tradeoffs to PTR as embodied in the HWA. The pharmaceutical industry operating under patents incentives augmented with PTR then becomes a unique setting in which to examine the confluence of historical and economic justifications for patents. The following chapters bring more detail to bear on the issue, leading to a model capable of supporting rigorous analysis of the tradeoffs in private and social benefits which come with patent term restoration.

_________________________

pharmaceuticals of using patents to incentivize both innovation (research) and clinical trials (development).
Chapter 2

The Economics of Patents

Never tell people how to do things. Tell them what to do and they will surprise you with their ingenuity.

Gen. George S. Patton

The goal of directing economic inquiry toward the patent narrative is to define the degree of protection for creators of new knowledge to receive. This statement presupposes of course that the degree is some positive quantity with the potential to be a net positive over any downsides. While this research also takes that assumption, it bears noting at the outset that there exist existential questions.¹

Taking their existence as given is obviously consistent with the real-world landscape. The legal reality quickly lends a corollary assumption in the form of universal enforceability of patent protection. As Clarke and Kohler (p. 19, 2005) introduce the issue,

¹We may again reference Edith Penrose’s (p. 40, 1951) classic statement, infra at p. 3. More recently, Grubb (p. 54, 2005), stated that [i]t is not a self-evident proposition that a strong patent system such as now exists in industrialized countries is in the public interest in these countries, still less that such a system is good for all countries in whatever stage of industrial development. The economic tools to carry out such a cost-benefit analysis simply do not exist.
“the decision as to whether a particular type of right should be recognized as a property right is dictated by policy reasons. Because property rights have [the] characteristic of enforceability against the world at large, they are dangerous things, capable of having adverse effects on people not even in contemplation when the rights were created. The justification for allowing enforcement of a right against one particular person is therefore not necessarily sufficient justification for allowing its enforcement against the world at large.”

With the expansion of patent protections at the end of the 20th century and their harmonization across countries under accession to the World Trade Organization’s TRIPs framework, patent owners are certainly able to enforce their claims uniformly (Kyle & McGahan, 2012). It is not entirely clear whether universally enforceable rights are social welfare maximizing in all situations. The problem is especially notably in pharmaceuticals, where exemptions for poverty and dire need generate much discussion.2 We limit the scope of the discussion here to the social welfare created under universally enforceable rights.

Under those caveats, this chapter introduces the economic literature’s rather large discussion on optimizing patent rights. The focus divides into two themes: optimizing length and scope of solitary patents and efficient distribution over cumulative innovation. Both areas draw from the balancing theme introduced via the history of the patent narrative (Chapter 1), where optimality starts with the amount of time necessary to guarantee a cumulative profit to cover the fixed costs to inventing and tradeoffs extend therefrom. To begin to formalize ideas, we can define this as the inventive threshold of a stand-alone innovation—one with no intellectual predecessors or offspring.

\[
\int_0^L \pi(t)e^{-rt}dt \geq C
\] (2.0.1)

The qualitative side of the inventive threshold is that it creates the break-even point for

---

2Notably, compulsory licenses. See, i.a., Gallini & Scotchmer, (2002); Tandon (1982).
firms. Projects whose expected markets evolve through time according to \( \pi(t) \) break even and pass the inventive threshold if the cumulative, discounted profits during the patent term \( L \) at least remunerate R&D-costs \( C \). Notably, costs are generally viewed as a fixed sum under the rationale that profiting commences after successful R&D. When viewing the patent as an incentive to invest in an R&D project, the firm can be said to require access to funding of magnitude \( C \) even if that expenditure will play out over the course of development.\(^3\)

Optimizing the length of the patent is straightforward in this simple setup. Longer grants benefit inventors by granting longer control over a profit flow. They also make viable a larger set of projects, given general assumptions on the distribution of costs to developing different projects at a given point of technological and scientific understanding. Patent breadth, however, is a more qualitative concept, requiring an understanding of the legal contours of patenting. The rhetorical definition of a patent’s breadth is by and large unquantifiable. Breadth plays a large role however in shaping an invention’s profit stream \( \pi(t) \). The reliability of the breadth’s definition, defined by its expected enforceability in court, should also play a role in the incentivization effect of a patent system. Taken together, length and breadth will define an inventor’s ability to profit from their invention.

Adding a final layer of complexity, considerations of inventive cumulativeness force a planner to consider the profit streams of closely related inventions together. The patent narrative’s intuitive appeal falters where the invention of knowledge does not have an immediate market value, or where that market value is small relative to follow-on development which attaches a significant marketable portion to the original information. Clearly the problem is minimized when the chain of invented information to the marketable invention is owned by a single inventor. Where it crosses owners, however, society leans on coordination mechanisms, such as licensing, to adjust ownership or tie profit flows together. While

\(^3\)This theme is also present in the Inducement Theory justification of patents, discussed infra at p.17, where ownership of intellectual property enables its use as collateral in acquiring development funds.
this dissertation does not consider the mechanics of licensing, we do discuss the social welfare improvements available from tying eventual profit streams together. We do not discuss licensing in depth because such coordination must have been completed prior to a drug entering the market. Furthermore, PTR is tied to a single patent so it may only serve as an incentive for the owner of the portfolio of patents covering the approved therapeutic.

Legal terminology runs throughout both patent scope and cumulative patenting sections and is therefore introduced first. Above the explanation of terms there is a goal of illustrating the patent system as an efficient system itself. There are problems with patents and there are problems with the administration of the patent system. As this research and economic inquiry in general takes the operation of the patent system as given, any suggestions of modification to the patent itself rests on the static efficiency of the system of administration of patents. The operation of the US Patent and Trademark Office (USPTO) is not generally thought of as efficient\footnote{See, for example, Bessen & Meurer (2008); Lemley, Douglas and Sampat (2006c); Shapiro (2004).}, and there is overlap between internal problems and subsequent legal and economic inefficiency. Such problems are discussed vis-à-vis their role in the emergent problems but otherwise administrative inefficiency is left for later research.

### 2.1 Hurdles and Claims

The issuance of a patent stands behind three qualitative hurdles given in 35 USC §101-103. They can be summarized as: 1) Novelty; 2) Non-Obviousness; 3) Usefulness. The European counterparts offer some conceptual clarity: newness; involving an inventive step; capable of industrial application. As patent applications are presumed valid\footnote{35 U.S.C. §282. Note criticisms of the presumption of validity in Lemley, Douglas, and Sampat (2006).} a proper review of the application and establishing that the new knowledge does, in fact, pass the hurdles, is the first step toward an efficient system.\footnote{Compare Lemley (2006c). Overwhelming the limited resources of patent examiners is a real concern. Patent examiners spend an average of only 18 hours on each examination, spread over three years (Lemley, Douglas, & Sampat, 2006). Lemley (2006c) argues that such a system grants too much power to weak innovations at the expense of the consideration a patent on a strong invention receives in court.} Understanding the hurdles and the role they play
in creating the patent-protected markets we see today is also the first step for economic research.

Novelty is the legal conception of the inventive spark itself. Invention occurred, according to pre-1952 US law, with a “flash of genius” (Grubb, 2005, p. 22). More formally, the inventive spark has created something novel if it adds to prior art. The prior art is a description of the state of knowledge in a specific area of expertise or the public domain.\(^7\)

The inventive spark is a dichotomous variable; there is no restriction on the magnitude of the step undertaken (Grubb, 2005, p. 64). Inventiveness is rather marked by whether someone working within the set knowledge consisting of prior art would stumble upon the concept in the normal course of work or research.\(^8\) We therefore prevent inventions which are in current existence as evidenced by prior art from patenting and those which can reasonably be said to be anticipated—caused to be obvious—by existing knowledge.\(^9\)

There is a legal fiction which serves as the comparative metric for questions of anticipation. This is a “person having ordinary skills in the art”, or PHOSITA.\(^10\) The term evolved from the times where mechanical inventions were the norm and the PHOSITA took the form of the “ordinary workman.” With the complexity of modern inventions the PHOSITA can be a team of highly qualified investigators and not merely a single person (Grubb, 2005, p. 63). The patent examiner charged with prosecuting—a term of art for the back-and-forth editing process between applicant and examiner—must gauge the prior art and obviousness.

\(^7\) Prior art is “a continuously expanding corpus of knowledge which has to be taken into account when assessing patentability (OECD, 2002, p. 24).” Laws of nature are not novel because they are discovered, not invented (Conley & Makowski, 2003). To illustrate that arguments still exist at these foundation levels, some argue that laws of nature should be patentable. See discussion in Rimmer (p. 119, 2008).

\(^8\) 35 U.S.C. §103

\(^9\) 35 U.S.C. §§101,102. But the conceptual difference between a discovery and an invention is less distinct in other contexts, such as computer software, information processing and especially biotechnology (Langinier & Moschini, 2002; Rimmer, 2008). The most vocal discussions center around gene patents, which have been ruled as inventions. They are currently patentable so long as they are “isolated and purified.” See, i.a. Scherer (2002) . There is recent contention on the issue, though in the form of legal definition and not moral arguments from non lawyers. This turns on clarifications of the isolated and purified standard, more clearly separating the two which had often been used interchangeably and used redundantly. See Association for Molecular Pathology (AMP) v. United States Patent and Trademark Office (USPTO), (Fed. Cir. 2011). Notably, one of the judges in the case was Alan D. Lourie, whose opinions and historical references in the Hatch-Waxman process are cited at length in Chapter 5.

situation. They do have some guidelines however.

The Federal Court has six criteria which determine the level of the skill of the art.\textsuperscript{11} These are: the educational level of the inventor; type of problems encountered in the art; prior art solutions; rapidity of innovation in the area; sophistication of technology; and educational level of active workers in the field.

Modern complexity has helped shift judicial focus from weighing “inventiveness” to “non-obviousness” under the PHOSITA perspective.\textsuperscript{12} Courts and examiners can now focus on proxy economic criteria and motivational indicators rather than only technical aspects of the invention at issue. Evaluating whether a creation is obvious or not relative to the prevailing prior art is arguably less subjective than establishing whether an applicant’s invention exhibits a flash of genius. It also prevents patenting of knowledge or equipment used by those in the field but owned by none. Otherwise, someone could argue that it was their flash-of-genius to patent and disclose local knowledge.

Objective factors are familiar to economists and include the commercial success of an applicant’s invention. Such evidence is considered dispositive of non-obviousness.\textsuperscript{13} Other dispositive factors can include: a long felt but unresolved need in the area; failed efforts of others; copying by others; praise for the invention; unexpected results; licenses established; industry acclamation; disbelief of experts; general skepticism; commercial acquiescence; and simultaneous development (Barr & Reisner, 2008, Sec. 5:3.7).

Patent examiners, however, are notably overloaded and are unlikely to be able to complete a thorough investigation into the particular situation of every application (Lemley et al., 2006). This is an important explanatory factor as to the amount of patent litigation. Prosecution of an application is, despite its length, relatively cursory. Litigation then serves to define a patent’s boundaries more clearly only when the need arises.\textsuperscript{14}


\textsuperscript{12}Graham v. John Deere Co. 383 U.S. 1 (1966)

\textsuperscript{13}Merek \& Co. v. Biocraft Labs., Inc., 874 F.2d 804, esp. 809 (Fed. Cir. 1989).

\textsuperscript{14}The patent search that one receives during their application carries very little legal weight should the
The efficiency of this situation is the subject of much legal writing, lamenting, and defense. Assuming a relatively functional patent office and patent litigation process however remains in the background when research, such as the current work, addresses the optimality of the patent itself. The optimality of a patent system is a larger undertaking.

Under that assumption, the first two hurdles are easily met with truly new information. The third hurdle of usefulness, however, constitutes the largest obstacle. Illustratively, the hurdle is known in European patenting as “industrial application.” Here we have a protection against patenting both pure knowledge and trivial widgets. This is especially important to remember in the context of pharmaceuticals. Many new chemicals can be and are synthesized in laboratories every year. Only a few are amount to anything more than pure academic interest, however. Even fewer make the leap from theoretical interest into practical interest that could be described as industrially applicable (Grubb, 2005, p. 212).

In the same vein, Intermediate compounds, useful only in their application to creating other compounds, are patentable under the logic that intermediaries of industrially applicable end products are useful (Grubb, 2005, p. 212). This final hurdle, which provides the broadest check against frivolous patenting, is also the closest to requiring an applicant to create a social benefit worthy of earning patent protection.

There is, however, no explicit requirement to create social welfare. Nevertheless, even unmarketable patents create new welfare as long as the patenting process results in official disclosure to society. Proving that something is new, novel, and applicable comes out through the prosecution of the application. In most of the world, the finalized application is then disclosed via publication. In the U.S., the disclosure occurs eighteen months or less after initial filing (Langinier & Moschini, 2002, p. 5). Disclosure has a threshold requirement too; an application must contain enough information that a sufficiently capable individual patent end up in court (Bessen & Meurer, 2008, p. 55). There are, however, specialized patent search firms who seek to protect clients from infringement and aid in patent preparation (Grubb, 2005, p. 371). There is some question as to whether firms do a search at all, however (Cockburn & Henderson, 2003).

15For an introduction, see Bessen & Meurer (2008).
could replicate the new invention. This known as enablement.\footnote{35 U.S.C. §112} Full disclosure is enough to enable a PHOSITA to replicate the invention. Of course, readers of the disclosure cannot utilize this knowledge to any marketable advantage. The access to new information and the enablement of replication is, however, seen as a catalyst for the next round of innovation.

The catalyst of disclosure is in marked contrast to another form of intellectual property protection: trade secrets. A full discussion of trade secret law is beyond the scope of this essay though.\footnote{For an introduction, see Friedman (2008,1991).} Trade secrets, while creating assertable rights when information is unjustly or improperly disclosed, provide no protection against eventual literal infringement—creating an exact copy—or equivalent infringement—a substantially equivalent innovation—where the competitor’s efforts are their own (Corbitt, 2008, p. 398).\footnote{Technically, there is no “trade secret law” sub-field akin to “patent law.” A violation of trade secret is a claim of breached contract or trespass and hence only prevent against conduct deemed “improper” through other areas of law (Risch, 2007).} Trade secrets are most effective in industries where reverse engineering would not pose a significant problem (Y. Friedman, 2008). As pharmaceuticals have traditionally been easy to reverse engineer, trade secrets offer little to no protection. The situation deserves a passing note, however, with the increasing complexity of biologics development, a situation which may require subsequent research to address strategic overlap between secrets and patents.\footnote{There is crossover here with active literature, both legal and economic, on the patentability of human genes. These instructive codes and snippets of DNA are themselves patentable. Usually, pieces of information inherent to the natural world are not patentable. Discoveries, in a pure sense of the word, are not patentable. They must be “isolated and purified” such that they are novel, and do not appear in the applied-for form in nature (Goldstein & Golod, 2002). There are many critics of this, including Goldstein and Golod (2002). Note Parke-Davis & Co v. H.K. Mulford & Co. 196 F.496 (2nd. Cir. 1912) in establishing a valid patent for adrenalin which was purified of inert gland tissue; in re Berstrom, 427 F.2d 1394 (C.C.P.A. 1970) for establishing that a purified prostaglandin extract from the prostate did not exist in nature and could therefore receive a patent; in re Bergy, 563 F.2d 1031 (C.C.P.A. 1977), which established that a biologically pure microbial structure could be patentable; in re Kratz, 592 F.2d 1169 (C.C.P.A. 1979) which found that substantially pure form of strawberry flavor could be patented; and finally the blackletter case Diamond v. Chakrabarty, 447 U.S. 303 (1980), in which the Supreme Court ruled that a human-made microorganism which did not exist in nature could be patented. See also Rimmer (p. 27-46, 2008). Notable to this progression is also the failed attempts to patent DNA which was not substantially isolated or purified from its natural state, such as in Ex Parte D, 27 U.S.P.Q.2d 1067 (1993). The case law shows that applicants should define themselves what they believe is the way in which their genetic discovery is purified from its natural state, as well as to prove that their discovery is not identical to any naturally occurring DNA (Goldstein & Golod, 2002) Above the novelty requirements specific to genetic patents, there is also a rather}
Most important amongst information in a patent are the claims. When we refer to application prosecution, we refer to the revising, editing and deleting of claims. What an applicant claims in their application is the substance of their novel, non-obvious, and useful. Claims therefore are what eventually surpasses the three hurdles. They also define the new body of prior art and become the legal boundaries of the protected intellectual property, boundaries which will simultaneously determine the size and shape of protection afforded in the market.

A patent's claims have a unique peculiarity, laying out as they do the feelings of the inventor as to what he or she contributed to society. When a patent is said to be "granted", it is control over these intellectual claims that is granted. Claims therefore become very nuanced as an inventor and their representatives work to stake out as much territory as possible. The examiner pushes back, as they are charged with limiting the grant to the exact contours of what was contributed to society.

The term scope is important term in both legal and economics literature on patents, and describes market area of a potential monopoly (OECD, 2002, p. 24). Claims, along with backward citations—references to existing inventions and prior art—locate the newness of the invention for the patent examiner (Hegde, Mowery, & Graham, 2007, p. 14). The inventor is the first to postulate what this newness-space consists of, citing the survey of the inventive landscape as they understand it. The examiner is then free to add their own information and updates the location described by innovators. This back-and-forth

strict bar set to establish the utility of discovery (Goldstein & Golod, 2002). See Goldstein and Golod (2002) and Genentech Inc. v. Chiron Corp, 112 F.3d 495 (Fed. Cir. 1997). It is also a bit problematic, in terms of science, as single purified genes may have more than one "utility" in the human organism; more than one purpose. While whatever utility is discovered in the laboratory is certainly one—whether that is coding for a certain protein or associated as a marker for a certain disease—it may be connected to other life processes (Dutfield, 2003, p. 163). Simply stated, genes are not independent factors. This fact leads to criticisms that patents could block later discoveries on other utility of that gene in the human. While this is possible, it is not a novel problem in patenting, as new "uses" for machines and processes could also be discovered after the original patent. In fact, this is part of the utility of the patent system as a whole—to encourage discovery of associated utility to disclosed information. Finally, we note an emerging concern that biologics are complex enough to fail enablement (Mandel, 2006). This is an ironic implication of current patent holders' claims on the regulatory side of the industry that competitors could not replicate their product safely. See Section ??
synthesizes the disparate information sets of examiner and inventor. What emerges is a territorial claim which does not overlap with others' claims but does establish the innovator's property. *scope*

From the introduction, we see that the market contours enabled by the new property will play a central role in the optimality of a patent. The scope defined by the claims however defies quantitative assessment. The physical analogy for intellectual claims, claims to a space of land, does not generalize entirely.\(^{20}\) Assessing whether a neighbor is trespassing on one's land is straightforward. The boundaries of intellectual trespass, composed only with words, are much more malleable.

There is some functionality to the fuzziness. Claims must be adaptable to encompass the ever expanding creativity of the human mind. More concretely, claim construction must be able define physical product or a process.\(^{21}\) For either type of inventive activity, claims will embody the prior art of the industry in which they are operating. When ideas find application and extension in other industries, malleability of rhetorical description creates questions of overlap that must be litigated in court.

The court are however specially capable in deciding these matters. A more pressing concern arguably lies in the misalignment of incentives for claim construction. The potential for larger intellectual areas to translates into larger protected markets inclines the inventor, who has the most accurate information as to the novelty and usefulness of the invention, towards the construction of vague claims. The examiner will of course attempt to clarify the claims. Examiner resources are, however, limited and attention is separated among several tasks.

\(^{20}\)See Kitch (1977). Ibsen (p. 186, 2012) notes however that there must be a distinction between "property" and "intellectual property" in that if "intellectual property was no different from physical property, there were no logical grounds for limiting their duration or legal status or even against making patents hereditary."

\(^{21}\)A product patent claims the functions the product serves. Process patents cover the methods and means of manufacturing items. Process patents become an umbrella for the products available from that process. Under the TRIPS amendments to the GATT, the treaty which provides for identical intellectual property protection across its signatory countries, including the U.S, process claims must be available and cover the direct product of the process (Grubb, 2005, p. 17).
Research shows that the average number of claims per patent application has increased, rising from 9.94 in 1970 to 14.87 in 1990 (Allison & Lemley, 2002). Of course, multiple claims as well as overly broad claims which pass the patent examiner can be scaled back by the stricter PHOSITA tests of a Court. The momentum however remains in the favor of the applicant across the board (Bessen & Meurer, 2008, p. 70). Starting at the largest and most vague point increases the likelihood of garnering an unduly large claim while sapping examiner time and resources. Broader claims also increase the opportunities for the patentor to contend that infringement has occurred (Langinier & Moschini, 2002, p. 9). Above the offensive aspect of non-trivial damage awards from infringement, the prospect of a suit can serve as deterrent against computers starting work in a given patent's area.

A limit to the applicant's behavior is that vague claims are more difficult to defend in court. An efficient level of ambiguity should therefore exist and serve as target construction for applicants. That level may be rather high however, given that only 5.8% of patent claims are invalidated on a basis of indefinite construction (Allison & Lemley, 1998). In pharmaceuticals, and especially with biologic drugs, there is however a recent, opposing concern that narrowing claim interpretation carries the momentum. This implies a growing ability for competitors to circumvent sector-specific claims (Holman, 2009). Assuming that the costs of such litigation were not themselves prohibitive, there could be a bounty set on testing innovator's claims.

Refraining from any empirical judgement, this brief introduction serves first as an introduction to patenting as a legal concept and second as an illustration of the difficulties in

---


23Some of the impetus for expanding claims come from Festo cases, (Festo Corporation v. Shoketsu Kinzoku KK et al 56 USPQ2d 1865 (Fed Cir. 2000 and 62USPQ2d 1705 (Sup. Ct. 2002), which, by their negative treatment of patent amendments in creating protection from infringement, pushed patent attorneys to write more claims instead of relying on later amendments (Grubb, 2005, p. 424). Prior to Festo, patent amendments had been used to later add claims to patent applications as new information and new legitimate claims came to light during development, information that did not exist at the time of the first filing but is also not great enough to warrant a separate or new patent application.
the implementation of a patent system. The imagery of incentivized claim construction sits awkwardly against any implied imagery of a well-defined object arriving before the patent examiner for assessment. The "Claim Game" (Bessen & Meurer, 2008), is the first hint that there are more issues to solve in pursuit of efficient policy than just defining an optimal patent term.

In the main, however, the patent system has proved remarkably adaptable. Computer software and gene patents may continue to stretch the application of the patent system's hurdles and legal rubrics but it continues to accommodate. Despite numerous hiccups there is yet no emergent phenomenon so large as definitively tilt Penrose's dilemma which opened the chapter. More germane to coming analysis, questions of efficient patent term restoration can be seen as a sub-question within the evaluation of a patent's total optimality. With this in mind, the following section surveys the economics literature surrounding optimal patent incentives to glean the dynamics necessary to model the subfield of patent term restoration.

2.2 Patents in Economic Theory

The economic investigations toward optimal patents is mostly easily introduced by extending outward from the break-even, remunerative threshold concept. Following Langinier and Moschini (2002), consider an inverse demand function, $D(p)$, for a newly invented good. There is a sunk cost, $F$, required up-front to develop an idea into the marketable product that will generate the demand relationship $D(p)$. The total new welfare created $\omega$ may change over time but the figure below presents a snapshot of $\omega(t)$. The variable costs of manufacturing are given as $c$ and are generally considered to be 0 in models of small-molecule production, owing to their small cost relative to R&D expenditure.

If the total surplus created is larger than the monetary costs of development, $F$, then the development is socially efficient. Facing sunk cost, $F$, no firm is willing to undertake

\footnote{A textbook treatment is Scotchmer (2004). Also Langinier and Moschini (p. 2-3, 2002)}

\footnote{This requirement is analogous to a production innovation which lowers per-unit costs enough to cover}
the development on themselves if they expect competitive forces to drive profits to zero. That is, even though the invention’s expenditure is socially efficient, the individual firm would not recoup their up-front expenditure and it is therefore not rationale to invest in the R&D project.

The text public good problem emerges now in the context of knowledge creation: a market without a lighthouse and an iPhone but market incentives for an individual to undertake either project are lacking. Unlike lighthouses, taxation for invention carries additional problems. Notably, taxing a populace for traditional public goods is assured to provide those public goods; there is no uncertainty in the building of a lighthouse. There is however much uncertainty in the invention of a computer or pharmaceutical. While there are benefits to an innovation approach which incorporates tax-based prizes and rewards (Wright, 1983), they necessarily rely on identifying a pressing problem. That is, society can its own expense at the current level of output (The Nordhaus/Scherer model) discussed in Section 2.2.1
identify a need and the outcome of addressing the need is well definable.

Certain issues do emerge as socially pressing and sufficiently definable to garner this attention. Notably, pharmaceutical development projects are often among them in that the need or detriment is well defined through a deviation from a healthy state and therefore the benefit is also clear. The larger current of innovation flows from undirected development. The benefits of most inventions are largely unknown during the development process and, furthermore, would be difficult to reach a political consensus on as to funding given the heterogeneous nature of individual’s utility. At the very least, the information needed for a central authority to come to locate projects and then come to a decision is non-negligible.

Offering a patent, on the other hand, requires minimal information and limited financial outlays on the part of the state. As we saw above, there are non-negligible management issues within a patent office and with the prosecution of application. Internal efficiency notwithstanding, leaving R&D decisions in the hands of those with the most precise understanding of the path from investment to product engages more market principles than any taxation-based award system. We proceed on the assumption that the distribution of knowledge in society finds the efficiency of the patent system to be net positive.

Referencing Figure 2.2.1, the patent allows a newly-protected monopolist to recoup R&D costs by using the monopolist’s pricing rule, setting $P_{\text{mono}}$ and quantity at $Q_{\text{mono}}$. Classically, the monopolist converts some consumer surplus into producer surplus, denoted by the profits in area $\phi$ at a given time. If $\phi \geq F$, development is undertaken rationally. Society benefits from the surplus area $S$ during the monopoly period as well while regaining the entire surplus should the patent expire. The social loss during the patent period is the dead weight loss (DWL) illustrated by triangle $l$. Here, $r$ is a discount rate attached to a discounting process $\delta(t;r)$, often simplified by assuming exponential discounting: $e^{-rt}$.

---

26The Bill and Melinda Gates Foundation is a particularly visual example. Although funded by donations and its endowment rather than taxes the concept is the same.

27Information-heavy options for innovation incentives include prizes and structured R&D races, as discussed in Denicoló (2000,1996).
Given a product costing $F$ to invent, and for the moment ignoring any secondary effects or inspiration which that first patent grants to other inventions, the optimal patent life span would simply solve the threshold question: \[ \int_{0}^{T} \pi e^{-rt} \, dt = F. \]

The economics modeling cited as the beginning of modern patent investigations shares this spirit but found more concrete expression by focusing on innovations as production costs reducers. The original work by William Nordhaus (1969)\(^{28}\) and the geometric interpretations which Scherer (1972) introduced, place the motivation for R&D on desire to raise productivity (Scherer, 1972, p. 422).

### 2.2.1 Patents for Productivity Increases

This section reviews Scherer (1972) in his abbreviated geometric interpretations of Norhadus (1969). The idea is straightforward; the level of research done, or investment therein, gives rise to an invention possibility function which traces the relationship between investment levels and per-unit reductions in production costs. The more investment, the more cost-reducing inventions society receives (Scherer, 1972, p. 422). The choices of the firm are naturally bounded by the underlying market demand for the established product.

The idea is captured in Figure 2.2.2. Where initial competitive conditions would compel a market price of $P_0$ for the producer's goods, securing a patent on cost-reducing technology allows for a lower price, $P_1$. The price competition forces high-priced competitors out of the market. The monopolist then produces the entire competitive output at the original price and collects the monopoly rent represented by the solid box. Equivalently, the patentee could license the cost-saving technology to the existing competitors and extract the same rent (Scherer, 1972, p. 422).

\(^{28}\)But see also references in Fritz Machlup, *An Economic Review of the Patent System*, Study No. 15, U.S. Senate, Committee on the Judiciary, Subcommittee on Patents, Trademarks and Copyrights, 85th Cong., 2d sess. Washington 1958; for comments on the economic inquiry of the patent system extending back over 100 years.
Innovation does not create a different product but rather enables the existing product to be produced at lower cost. The existing competitors remain in the picture and limit the otherwise classic behavior for a monopolist to reduce output and increase price. Competitors can stay in the market, producing some of the market demand using the older technology with higher costs. As long as demand is not too elastic around the original price point the monopolist will continue pricing at the competitive price and quantity combination (Scherer, 1972, p. 423). When the patent ends, all producers may adopt the technology and the cost of the product is now lowered for this time forward.

These results however only hold for small innovations which do not effect unit costs “too much.” Dramatic innovations could reduce unit costs by more significant quantities, say, down to $P_2$ in Figure 2.2.2. In the case of drastic inventions, the innovator would then find it optimal to reduce cost and expand output past the previous competitive output levels. (Nordhaus, 1969). The super-technology makes limit pricing a credible threat. As long as
the R&D steps are small however the annual monopoly rents grow linearly as a function of the percentage of unit cost reduction, $B$, given the government’s imposed patenting time limit, $T$. Graphed along with the Invention Possibility Function (IPF), under Scherer’s (1972) assumptions of initial regions of increasing returns to research followed by decreasing returns, the linear function $Q(B, T)$ of quasi-rents reveals the regions of net gains between itself and the “cost of invention” IPF function (Scherer, 1972, p. 423).

![Figure 2.2.3: Scherer's (1972) Conceptualization of Patenting's Effect](image)

Any decreases in the patent length, $T$, will rotate the IPF function inwards around the origin, decreasing the optimal research investment by shrinking the difference between invention possibilities and quasi-rents.$^{29}$ Increases in patent life similarly push optimal

$^{29}$See Scherer (1972) p. 423, figure 3
research investment outward. As the government increases patent lengths, R&D investment increases as cost reductions increase. This comes though at a social cost equal to the area of the triangular area in Figure 2.2.2 and is compounded over the lifetime of the patent. The cost of the patent becomes the sunk R&D costs plus the compounded welfare loss. The policy maker therefore simply balances costs and benefits through their choice of T.

The Nordhaus/Scherer model is notable both in its historical role but also in establishing a tractable social balancing situation that is not predicated on the creation of a wholly new market. The simplification of innovations as cost reductions avoids the larger assumptions which drive the results of future patent literature. Those complications are necessary however given the realities of the innovation we see and the inability of this model to capture drastic improvements. Figure 2.2.4 illustrates the change in the quasi-rent function when one considers much larger corresponding production cost reductions. These are the inventions which bring both cost reductions and expansion of output at competition-limiting low prices (P₂).

Now one has a new region of increasing quasi-rents relative to a leveling of the cost-to-invent function. Graphically we see that the horizontal distances increase quickly and, as research expenditures are determined by the maximum horizontal distance between lines, optimal research expenditure and hence quasi-rents explode.

Within these limitations, Scherer’s (1972) graphical representation makes Nordhaus’ (1969) three major conclusions clear. First, the larger the demand elasticity around the competitive and post-invention market prices, the larger the social losses. Second, steeper IPFs, capturing an easy invention, lead to shorter optimal patent choices. Finally, the more bowed the IPF is, the shorter the optimal patent life due to decreasing returns between induced cost reductions and increased patent life. (Scherer, 1972, p. 424-5).

30Chu (2009) predicts a weak effect for lengthening but a stronger effect on shortening when we look at pharmaceutical patents specifically.
Although this dissertation, and most research on pharmaceuticals, deals with predominantly with brand new and “creatively destructive” inventions, it would be too hasty to fence the results of these foundational papers out of the development of the work at hand. Here is a very general illustration of the potential for dichotomous desirability of patent policies between mild innovations and quantum-leap technologies. It is the first indication that one-size-fits-all patent policy is intrinsically hindered when turning a blind eye to the market in which the product will reside. The lack of a single efficient patent policy is a theme revisited throughout the economics literature.

\[\text{\textsuperscript{31}Scherer's (p. 426,1972, p. 369, 1970) creative categorization of patent policy's two separate effects, "stimulus" and "Lebensraum," is also notable. The stimulus effect is how the ex post promise of profits lures ex ante investment. Lebensraum requires that a "patent grant must persuade investors that competitive imitation will be deferred sufficiently long to make discounted quasi-rents exceed [research] outlays for at least some positive investment level (Scherer, 1972, p. 426)." Here is an early and explicit illustration in the economics literature of the patent as a necessarily remunerative device. The terms dovetail with current pharmaceutical commentary which uses such terms as "make hay" (Hollingshead & Jacoby, 2009) to describe generating adequate remuneration in the patent period.}\]
2.2.2 Two Levers: Time and Scope

The introductory comments as to the break-even threshold and now the Nordhaus/Scherer model seem to imply that efficient patents are not going to infinitely lived. These models however assume much: no competition in R&D phases; perfectly successful innovation efforts; and each R&D project results in a revenue-generating item, among other simplifications.

Economists are of course well trained in attacking assumptions. The assumptions of the social innovation process though seem to be particularly crucial to outcomes in the patent literature. There is rather little overlap between conclusions in papers which share similar assumptions as to the innovation progressions. DeBrock (1985) explored this explicitly. Therein he also criticizes background patent narrative or drawing assumptions from the reward theory of patents (DeBrock, 1985, p. 225). As we have seen, an assumption that the size of the reward is somehow related to the market value after invention has been a driving narrative for some time. DeBrock (p. 241, 1985) reminds modelers that "[...] rivalry has effects on the R&Dprocess; just as clearly, the R&Dprocess affects rivalry." The rivalry process is how inventions play off one another; where competition is fierce, innovation may proceed at a rapid rate were it not for a patent of fixed length. Replacing innovations at a rate prescribed by the initial market value of a patent forces improvements to wait in the wings.

Remuneration does not have to proceed on a fixed schedule to serve as a successful incentive however. When one needs to refill a pitcher, there are two variables to be adjusted: the length of time to leave the tap open and degree to which they open the tap.32 In the economics literature, questions of a patent's duration are separated from investigations of a patent's scope: the degree to which the tap is opened. The additional dimension aides policy in tune the patent/market bargain under expanded assumptions.

---

32This analogy is common among patent papers. Inter alia, Gilbert and Shapiro (1990), with their visualization of breadth as a flow rate.
Where a patent length is simply a duration, scope describes the degree of protection from non-infringing competition. Scope therefore encompasses some measure of similarity between products. In optimal length calculations, we are only concerned with the hurdles of invention and the claims stated in the application. A separate product which does not step into another's claims does not infringe the first’s patent and may themselves garner a patent and profit from their invention. This is true even when the second comer’s product changes the market of the first. Were a patent examiner to take this into consideration explicitly, the stated patent length which lent an expectation of solving the remunerative threshold upon entry could suddenly fall below the threshold non-infringing competitors entered. On the other hand, the problem of broadly stated claims introduced earlier could combine with a static patent life to over-remunerate the invention. The degree to which an inventor can push the market boundaries of invention is not only important for the profit of the firm but for freeing society’s choice in patent length.

Often, the term scope is implicitly broken into breadth and height. Breadth carries an implication of horizontal diversification. A patent’s height is then the degree to which a follower must improve upon—or “leap frog”—the original in order to garner its own patent. Taken together, “breadth” and “height” of patents dictate the competitive boundaries available for profit. But while the length of a patent is easy to define in years and months, patent scope, as an interpretation of the underlying claims based on economic rationale, must be defined by words and the law.

The term *scope* tends to encompass both a patent’s *height* and its *breadth*. The greater the breadth, the “less similar” a competing product can be to the original; the greater the height, the more superior or displacing an improved version must be. The concept of height falls under “leading breadth” as it protects against “future” innovations while breadth protects against contemporaneous competitors.33 Note that these are economic terms and

---

33Langinier and Moschini (p. 10, 2002), in their literature review, credit van Dijk (1996) with the term “height” and O’Donoghue, Scotchmer, and Thisse (1998) with the term “leading breadth.”
do not reference any current reality of patent prosecution which, as the introduction shows, resides on an analysis of stated claims in a patent relative to others’ claims. The terms do however exist in legal literature but are not always reflective of a particular usage from the economics literature.

The rhetoric also clarifies what it means to infringe a patent. A patent infringes another when it comes within the length or scope of the original’s protection as defined by the listed claims during the patent’s term. The case of direct copying is straightforward; one cannot profit from the patented claims of another during the patent’s term. The competitors and future improvements which may or may not infringe the scope of claims is another thing entirely. It is often difficult to define the characteristics which make one products compete with each other. Perloff et al. (1996), for example, in search of a quantitative handle adapt spatial games with quality improvement choices to the patent context. There are many other handles however and we discuss the major contributions below. To foreshadow the outcome, the literature agrees to disagree, and each option seems to have its place and usage (Erkal, 2005).

When considering a patent’s length, we established that longer patents are worse for welfare, assuming that the patent length is sufficient to call for the the invention. When considering the scope of a patent grant, broad patents inflict the largest costs on society stemming from the large monopoly influence. Narrow patents are less distortionary with respect to welfare as closer substitute goods may appear. Gilbert and Shapiro (1990) look to tune length, $T$, and breadth to balance the incentive and welfare tradeoffs.

Breadth is described as a flow rate of profits, $\pi$ (Gilbert & Shapiro, 1990, p. 107). Greater values indicate higher flow rates to the patent holder during the life of the patent and thereby a faster refilling of a given R&D expenditure. A social planner maximizes social welfare, $W$, by choosing $T$ and $\pi$, subject to breaking the adequate reward barrier, $V$. The flow rate affects social welfare inversely, $W'(\pi) < 0$ and after patent expiration available welfare increases to its competitive equilibrium, $\bar{W} = W(\bar{\pi})$. 
Given these definitions, we express total social welfare as

$$\Omega(T, \pi) = \int_0^T W(\pi) e^{-rt} dt + \int_T^\infty \dot{W} e^{-rt} dt$$  \hspace{1cm} (2.2.1)$$

which is maximized subject to the present value of the patentee’s profits.

$$V(T, \pi) = \int_0^T \pi e^{-rt} dt + \int_T^\infty \dot{\pi} e^{-rt} dt$$  \hspace{1cm} (2.2.2)$$

If changing the flow-rate of profits is increasing costly in terms of social welfare, $W^{''}(\pi) < 0$, then infinitely-lived patents are the efficient policy that emerge from the maximization.\textsuperscript{34}

This obviously sits in stark contrast to finite logic above (Gilbert & Shapiro, 1990, p. 108). Given the choice, elongating the patent life $T$ is welfare enhancing relative to damages brought through increases in the flow rate $\pi$.\textsuperscript{35}

A dynamic problem emerges if we want to add incentives for continued innovation however. The concept of prospecting for innovations, expressed early in Kitch’s (1977), envisions the problem of an owner “squatting” on their mining claims simply because they own them. As property claims are eternal, analogy brings the same concerns to Gilbert and Shapiro’s model, though the authors readily note the simplicity of their construction.\textsuperscript{36}

Gallini’s (1992) paper tackles such an extension by finding what triggers firms to “invent around” patents. That is, a follower firm has some incentive to invest in costly innovation in order to work-around an obstructive innovation—or squatter—in the intellectual space. That competitive attack addresses the height of the patent’s claimed and protected scope. Extending patent life hurts the first inventor by inducing more aggressive search for work-arounds. But inducing companies to re-invent the wheel, as it were, by tying up established knowledge in patents for too long is, however, socially costly (Kotlikoff, 2008, p. 15). Under these circumstances, Gallini (1992) shows that when patent length is the only policy

\textsuperscript{34}Proof omitted. See Gilbert & Shapiro (p. 108, 1990).

\textsuperscript{35}Also note Tandon (1982) with a similar result motivated by compulsory licensing requirements.

\textsuperscript{36}E.g., on p. 112.
instrument, short patents are best to discourage imitation.\textsuperscript{37}

Gallini's (1992) paper begins the march toward adding more competitive elements into the patent narrative. These enrich the discussion but require more assumptions as to the nature of intellectual competition. These are relatively stronger assumptions than those governing a profit maximizing firm investing in R&D based on expectations of remuneration under the patent term or a social planner maximizing welfare with that patent term.

For example, Klemperer's (1990) employs a spatial differentiation approach where a broader patent carves out a larger chunk of the available consumers. The binary choices which these consumers then face drives a deadweight loss. Increased patent scope can forces customers to choose the original or an alternative products within the same product class, thus imparting a type of deadweight loss through substituting away from the preferred product offered by the monopolist. Imitation here is assumed costless in order to focus on the length and breadth tradeoffs. The substitution is modeled as a transport/switching cost, increasing in the distance the customer locates from the desired, but patented, monopoly product. The marginal consumer can leave the market but suffer a deadweight loss of leaving the product class entirely (P. Klemperer, 1990, p. 114-5). The generality permits conditions where either (1) broad & short patents (2) long & narrow patents optimize the patent/market balance.

In Klemperer's (1990) results turn on the specification of switching costs.\textsuperscript{38} He summarizes:

"if all consumers have the same per-unit transport costs of substituting to a less-preferred variety of the product, then infinitely lived, narrow patents are optimal.[...]. On the other hand, if for each consumer, the value of consuming

\textsuperscript{37} Licensing is not discussed in this work, but can serve a coordinating role between early and later inventors such that not all remuneration must derive from a single patent. The downside to powerful, broad patents can be addressed by compulsory licences (C. Shapiro, 2001; Tandon, 1982). More complicated cases where the compulsory aspect hurts incentives to innovate can be handled by open licenses with limits on fees (von Hippel, 1994, p. 52).

\textsuperscript{38}Klemperer (1988) also models in the pharmaceutical market the idea of switching costs, transport taking the form of moving from branded to generics."
the preferred variety exceeds the value of consuming no variety of the product by the same monetary amount, then short-lived patents that are as wide as possible are optimal (P. Klemperer, 1990, p. 115).”

Denicolò (1996) finds markets where either long & narrow or short & broad patents are optimal. The objective in this model is to minimize the ratio of social loss to innovation incentives for inventors. Fixing a social value to R&D in this situation is restrictive, so Denicolò relaxes that assumption by allowing for patent races. Races allow for the reward to innovation to be appropriated by the single winning party. When two or more parties compete for that same slice of pie, Denicolò (1996) finds that optimal patent conditions depend subtly on the assumed relationships between “social welfare and post-innovation profits, on the one hand, and the breadth of the patent, on the other hand (Denicolò, 1996, p. 263).” By subtle, Denicoló (1996) is referring to the fact that the conditions necessary for solutions extend to assumptions on second-derivatives. The word “assumed” is also importantly illustrative; economic theory gives researchers little traction in restricting the dynamics of functions used (Denicolò, 1996, p. 263). The general finding though is that reducing patent breadth increases competition among innovators. The increased competition however brings its own social costs such as duplication and inefficient production (Denicolò, 1996, p. 263). Then lower the efficiency of competition in the product market becomes, the benefits to broader and shorter patents emerge (Denicolò, 1996, p. 264).

Considering the ease or difficulty of inventive activities, La Manna (1992) found that infinitely long patents—a la Klemperer (1990)—can be optimal if one restricts the height enough. This is especially the case with “easy” innovations. La Manna (1992) finds that social welfare is best served by setting high patent standards instead tuning patent lengths. Such high patent standards however hurt incentives for the patent holder. Work from van Dijk (1996) finds the counterintuitive result that a protective measure can actually hurt the protected person. This stems from the credibility that the height hurdle gives to commitments in research programs. In van Dijk’s (1996) duopoly game, patenting reveals
the level of the inventive activity and thus the level an improvement necessary achieve a patent far enough ahead of the original patent to avoid infringement of their claims. In some specifications of the height—specifically a medium height—the hurdle allows the competitor to credibly choose to pursue higher innovations on the original patent knowing the innovator has sunk themselves in to a path too. When the height hurdle is however increased, the intuitive result that more protection benefits the patentee returns (Dijk, 1996).

The potential for opposing results illustrated between these works starts to reveal the sensitivity of patent models to relatively un-contentious assumptions. Research generally moves to empirical study in these situations. For example, Guell (1997) measures size of the deadweight loss in pharmaceutical switching and attempts to match that to a structural model. Guell (1997) estimates a range from $3 to $30 billion in the (broadly defined) pharmaceutical industry, amounting to inefficiency equal to 60% of sales. Narrowing the focus to a class of therapeutics finds $5 billion in foreign surplus.

Determining the degree to which these narratives on market conditions are valid is not at issue in this research however. A focus on a certain industry, such as pharmaceuticals, would surely help separate valid modeling assumptions though. Such is the raison d'être for this section., after all. The goal of introducing and cataloging the dynamics addressed by the patent debate thus far continues with the illuminating literature on the cumulativeness of inventions, where research addresses efficient distribution of aggregate rewards among the component, inspirational pieces.

2.2.3 Cumulative Innovation

The imagery of luminaries all standing on the shoulders of their predecessors is often employed to express the cumulativeness of innovation.\textsuperscript{39} Visualizing the innovative process as

\textsuperscript{39}The imagery, most prominently introduced to the economics literature by Scotchmer (1991), is often attributed to a speech by Sir Isaac Newton. Newton's famous quote however may have been a veiled jab at a short colleague rather than an insightful commentary on the evolutionary nature of invention (Bryson, 2004).
a ladder of connected success recognizes the fact that inventions, more often than not, build off one another.

We have already discussed the disclosure requirement of the patent, a requirement based on this logic. While an arguably important dynamic, connecting innovations in a chain of ideas and markets fairly complicates the simple picture with which the chapter started. Specifically, if each invention is more than an island unto themselves, drawing boundaries around the market which the invention brought forth, the market from which we assumed the invention would draw remunerative profits, loses its definition.

This section uses “cumulativeness” as an umbrella term covering the literature on sequential innovation, such as Green and Scotchmer (1995), but also notes its use in literature focused on arranging complementary technologies, such as in Lerner and Tirole (2004). The dividing line is whether the cumulative invention requires the combination of several other inventions—assembly of ideas—or whether it’s a combination of inspirations.  

---

40Langinier (2006) also discusses this separation point.

41The intellectual-assembly problem is a type of cumulativeness which embodies similar dynamics to problems in land assembly literature. The defining paper in this section of the literature, from Green and Scotchmer (1995), focuses on the former dynamic.

We assume that first innovations catalyze followers. The first innovation, however, may not have much marketable value in and of itself. The sum impact of inventions one and two may by substantial but the share only available once invention two has entered the market. If invention one is also the more costly of inventions, as is often the case with foundational technology that is later worked into applied and more marketable forms, the market may under-incentivize invention one and then, should it still manage to be invented, over-reward the other. An optimal innovation policy in a cumulative world should therefore recognize backward protection, or lagging breadth (O’Donoghue, Scotchmer, & Thisse, 1998). Scotchmer (p. 322, 1996) that “the problem of dividing profit seems particularly acute when the entire commercial value is contained in the applications facilitated by the basic research, and when the basic innovation has no commercial value on its own.”

If path breaking inventions are the backbone of an evolving knowledge ladder, then it makes sense to protect those starting new ideas more than improved ideas. If the original inventor is also the best situated to bring improvements on the original idea to the market, then the greater patent protection does not damage society in the same way as larger protection causes for other situations.  

There must be times where protection out in front of inventions is therefore more beneficial. Green and Scotchmer’s (1995) benefit lies with detering too-rapid “leap-frogging” so that the follower cannot absorb profits on the line of similar invention that would have accrued to the original inventor. Leap-frogging is different than imitating in that it improves on the original significantly enough to avoid patent infringement but will displace the original idea in the market. The concept implies that the supped-up item would could not have been invented without the original’s first-steps. Leap-frogging is therefore socially beneficial in that society gets fast improvement on originals but has corrosive effects on the incentives to invest in early R&D.  

Examples where an invention kicked off a firestorm of inventive activity are easy to find. There are also
Scotchmer (1996) even finds scenarios where \textit{excluding} patents on second-generation is an optimal solution. The result hangs on the underlying type of invention, abilities of the parties and the existence of licensing opportunities to avoid stifling innovations (Scotchmer, 1996). Real-world studies have however found that the threat of infringement suits does not generally stifle first-generation innovation (Lemley, 2008; J. P. Walsh, Cohen, & Arora, 2003).\footnote{Lemley (p. 29, 2008) also notes instructively that marketable pharmaceutical products are not generally cumulative in the same sense as in other industries.}

Once again, there is a balancing act for the innovation policy. And also as with patent length, there is no consensus as to the optimal amount of forward and backward protection. Also similar to patent length results, the conclusions as to optimal policy under cumulativeness break down into categories. Denicolò (p. 263, 1996) identified that part of the problem lies with the inability of economic theory to place relevant restrictions on the concavity on the functions which drive the tradeoffs: social welfare, post-innovation profits, and patent breadth. Categorical conclusions are therefore to be expected.

There are two major outcomes: situations which benefit from forward protection and, by extension, the prospect theory of patents; and situations where forward protection stymies the evolutionary progress, implying net benefits in remunerating with wider and shorter protection. Mirroring the previous discussion on optimal length, the formal conclusions in this literature also rests heavily on what tradeoffs a modeler includes and how they specify how benefits accrue. As an illustrative example of both sides and a concrete summary of the dynamics which this research takes into the pharmaceutical sector, we examine Scotchmer stories of the confluence of many seemingly innocuous innovations into a torrential breakthrough though. The latter drives concerns with an emerging “patent thicket” (C. Shapiro, 2001) and an anti-commons effect from the over-ownership of knowledge units (Heller & Eisenberg, 1998). Specifically, Heller and Eisenberg (1998) and Merges and Nelson (1990) find support for the claim that forward protection stifles socially valuable second-generation products. The \textit{ex post} nature of the arguments for forward protection also finds criticism from legal circles under Professor Lemley’s work. Lemley (p. 131, 2004a) states that

\textit{“[t]he ex post justifications seem to provide economic support for the legions of new intellectual property owners who claim a moral entitlement to capture all possible value from “their” information—a view that scholars have derided as “if value, then right” (Lemley, 2004, p. 131).”}
Scotchmer & Green’s (1990) model focuses on incentives when an innovation is more valuable than the invention off of which it builds. The extreme case would be an invention which holds no intrinsic value but generates much value in the form of subsequent marketable invention. (Scotchmer, 2004). Whether there is initial value or not, the initial patent serves as disclosure to followers. Disclosure reveals information both on what research paths work and the probability for eventual successes. This is valuable information to the follower but comes at the expense of the inventors who search out a once-hidden path. The patent balance here has to compensate inventors for the externality of spreading the original ideas. Without adequate compensation the inventor finds incentives to strategically defer patenting to hold onto the idea, essentially hiding R&D signposts for followers, but secures an innovative head start. Society loses where the information remains private and undisclosed.

The concern then is that the sharing-externality induces firms not to patent anything, hoarding information to their advantage and depriving society of the effects of disclosure (Scotchmer & Green, 1990). Secrecy serves as ersatz protection by forming—or keeping—barriers to entry, forcing the follower to first fully reinvent the wheel and then improve upon it.46

The upside to patenting comes in as policy lowers the threshold of novelty for patenting. With a low height for protection to inventors it encourages disclosure of even small inventions punctually and rapidly. Momentum then can build behind an innovator who then routinely improves and keeps similar ideas from nibbling at their evolving market share (Scotchmer & Green, 1990; Matutes, Regibeau, & Rockett, 1996). The beachhead dynamic

46Trade secrets, were briefly mentioned above. Again, a trade secret is “an item of information—commonly a customer list, business plan, or manufacturing process—that has commercial value and that the firm possessing the information wants to conceal from its competitors in order to prevent them from duplicating it” (D. D. Friedman et al., 1991, p. 61). See also, §757 Restatement (First) of Torts. Trade secrets can be formulae, patterns, programs, devices, methods, or techniques and processes of economic value. (Uniform Trade Secrets Act §1.) For the fundamental treatment, see, inter alia, Lemley (2009), Landes and Posner (2003) and Bone (1998). For criticisms, see Risch (2007) and
forms a large part of Green & Scotchmer's (1990) conclusions, which lean toward lower novelty requirements. The low novelty requirements minimize the larger disadvantages in the cumulative innovation tradeoffs of boxing competition by putting more reward on those who patent small, early, and improve often.\footnote{What if the improvement cannot be reached by the original firm? Langinier and Moschini (p. 11, 2002) note the pertinent example of small biotechnology firms who do not have the capital or experience to expand pure research into FDA-approval oriented drug development.}

Formally, the paper models the interplay of two firms—an leader and a follower—competing in R&D. They conduct R&D under a given patent regime. Firm 1 is a leading firm who invents a product of quality \( x \). The product is patented and has a stand-alone (monopoly) market value of \( x \). This value cost the inventor \( c_1 \) to invent. Firm 2 gets inspiration from the first invention and innovates a next-step product of quality \( y \). This product has an intrinsic value \( y \), and cost \( c_2 \) to develop from Firm 1’s starting point. The quality and cost pairs are drawn from some distribution \((G, H)\). The draw occurs before Firm 2 decides to invest but after Firm 1 has sunk its costs into \( x \)’s development.

The \( y \) product will never enter the market without the presence of \( x \) as it relies on \( x \) for its foundation. That does not imply \( y \) inventions are somehow limited. Firm 2’s improvements can be extremely innovative over its precursor \( x \), a leap which is reflected in the profit of the second item. The model imposes novelty hurdle, \( y^* \), above which Firm 2 will not infringe Firm 1’s leading invention. Conceptually, as the \( y \)-quality increase, the original \( x \)-quality becomes a vanishingly small part of the total knowledge embodied by \( y \).

Infringement is a motivating factor for Firm 2 to invest in improvements. Infringement means that Firm 1 has legal power over Firm 2 and can prevent product \( y \) from entering the market.\footnote{From an economics perspective, it might make sense to have damages instead of injunction as a penalty. U.S. precedent however works within intellectual property law rather than considering any anticompetitive problems of injunctions (Lemley, 2009, p. 1033). The Supreme Court in \textit{eBay v. MercExchange}, 547 U.S. 388 (2006) moved slightly towards economic rationale in requiring plaintiffs to justify injunctions on a case-by-case basis.} This sets up the market’s competitiveness in R&D. Should product \( y \) prove non-infringing but still shares a degree of substitutable with \( x \) it enters the market as a
substitute and the two compete in a duopoly. In that case the competing firms would secure profits of $\pi^c_x - c_1$ and $\pi^c_y - c_2$, with the superscript $c$ denoting competitive pricing and subscripts denoting inventor and innovator. If Firm 2 enters and is ruled infringing, it must either leave or secure a license from the originator to produce.

There is room for bargaining based on the unknown outcome of infringement. The firms can bargain before Firm 2 enters and tests their quality improvement in court or they can wait until an infringement hearing occurs. The possibility that $y$ will not break the novelty threshold and will therefore infringe leaves Firm 2 in a poor ex post bargaining situation, with sunk costs and no market access and with Firm 1 holding all of the cards. They can however lower the probability of being ruled infringing by innovating at higher $y$-levels.

If the two firms can reach an agreement before entry, the firms split the eventual surplus and absorb their own R&D costs. Firm 1 wields bargaining power but cannot capture the full surplus, as the increase in total surplus from Firm 2’s entry gives the follower some leverage. The possibility of non-infringing entry without ex ante agreements also limits Firm 1’s stance, as they might end up in a duopoly situation without any legal recourse from their intellectual property.

In the end, Firm 2 can be held up by both choices, agreeing to or declining the license. Due to the uncertainty of the value of the innovation during the earlier licensing phase there is no guarantee of a positive profit for followers. Despite anti-commons concerns however there is also no way for the leader to secure the entire cumulative surplus. This is an important result, as patents are often implied to transfer disproportionate profit to a leading firm.49

While the details are left in the original paper, this overview introduces plausible dynamics of information spreading and tractable modeling techniques which finds times where patent life could be set to cover more the individual’s cost of R&D. If patent length, $T$, is tuned to cover costs $c_1 + c_2$ then the policy maker is failing to recognize the power of

49See Bessen (2004).
followers and is lowering the incentives for Firm 1 to initiate the innovation chain. There even is a chance of the inventor receiving a negative profit in their specification.

Bringing this result back into the canon of optimal patent life, the cumulativeness results in a recommendation that the more profit one can shift to the first inventor in a cumulative situation, the shorter patent life can be (Green & Scotchmer, 1995, p. 26). To an economist, this result likely emerges as unsurprising. The remuneration philosophy in the patent narrative was too simple even from the start to either prevent disincentives for investment between firms or to maximize social welfare. As Geroski (p. 426, 2005) summarized,

“...what is really puzzling about the intellectual property rights system is the way that it goes about preserving such incentives [to innovate]. For competition economists, the natural first step is to sum up all the investments made, allow for a bit of risk and compute a rate of return that the innovator ought to be allowed to earn on that investment. But, intellectual property rights regimes typically grant inventors monopoly rights for a fixed period of time regardless of their costs, or, for that matter, of the social value of their innovation. Further, these intellectual property rights place very few restrictions on the kinds of licensing provisions that inventors can impose on those who wish to take advantage of their work, allowing them, in principle, to earn phenomenal returns in some circumstances.”

It is at this stage then that the economics literature turns to pursue more and more realism in patenting, incorporating licensing and patent pools in the analysis (Langinier, 2005; Lerner & Tirole, 2004; Merges, 2001; C. Shapiro, 2001). Other lines of research include analysis of parallel protections like trade secrets, investing in barriers to entry, and outright collusion (Lemley, 2009; Erkal, 2005; Kong & Seldon, 2004). We introduce these briefly below primarily for their criticisms of the current patent system.
2.3 Licensing and Coordination

The broad rhetorical arguments on a patent's optimality from legal circles and the quantitative economic models explored above show that development hold-ups, offensive blocking, assembly problems, and bargaining conflicts are all realistic dynamics which change the incentives to innovate. In changing the incentives, the pursuit of efficiency with the patent system has more and more demands to balance. In the preceding section, we saw the necessity of licensing as a method for exchanging profits within a cumulatively-created market. Licensing does in fact play an important role in the competitive world of R&D and it occupies its own niche in the innovation policy literature. For the needs of this research, this section can focus on licensing alone. It is central in the pharmaceutical industry and many other coordination techniques share the same foundations.\footnote{See, i.a., Shapiro (2001).}

Preventing coordination failure in R&D has huge benefits. Some benefits were captured in the game theoretic results of Green & Scotchmer (1995). Outside economic estimation of surplus gained, coordination failure results in real-world litigation. Economic models tend to assume away the costs of using the legal system where they are not explicitly the focus of the investigation. In reality, intellectual property infringement litigation is long and expensive (Hope, 2008, p. 266). The legal realities do, in fact, shape the innovation processes we observe, notably in pharmaceuticals (Eisenberg, 2003). Especially for the sea of smaller start-ups who share the Big Pharma's industrial landscape alongside previous startups who have crossed the approval hurdle (Chapter 3), the damage resulting from infringement litigation can obliterate the assets of a entrant.\footnote{For an introduction to the industry structure, see Friedman (2008) and Chapter 3.} R&D, not just market entry a la Green & Scotchmer (1995), is a large risk for start-up companies whose entire portfolio revolves around a handful of intellectual assets (Hunt, 2006).

The American Intellectual Property Law Association estimated that patent litigation for each firm engaged in R&D ranges from $0.5 to $5 million (Hunt, 2006). There is almost
no limit however at the top end of litigation. For example, Cordis and Boston Scientific battled at a cost of $271 million while 3M and Johnson & Johnson burned $107 million. It is broad technology patents, like those more often owned by small companies working at the forefront of research, that are more likely to be litigated (Lerner, 1994). Though there is a wide variation in the probability of litigation across patent categories and the average threat of litigation on IPR is about 1% (Lanjouw & Schankerman, 2001, p. 131), the fact that litigation can become an existential problem will naturally change, at the very least, the bargaining scenarios described above.\footnote{A nearly-inelastic demand made royalties for end-products high enough to deter successful licensing in the early drug industry (Temin, 1979).}

The problem is even more acute in the pharmaceutical industry. Highly valuable drugs and therapeutics patents have a 25\% chance of facing litigation, far exceeding the 1\% average (Lanjouw & Schankerman, 2001). On its own, this is a scary landscape for cumulative innovations. More complicated though is the emergent behavior stemming from the repeated nature of the R&D patenting game. Expensive litigation initiated immediately to stop a relatively minor infringement can be used to set the precedent for future deterrence. Given the global reach and uniform content of IPR within WTO countries, patents are often litigated in different jurisdictions. An infringer stopped in one country, such as Europe, will likely settle out of court in other countries however (Grubb, 2005, p. 432). If a patentee has a global stake in their patent, they’ll often choose the US as their primary litigation forum, as the process there tends to set the tone for worldwide settlements (Bessen & Meurer, 2008, p. 5).\footnote{This phenomenon can also lead to forum-shopping for the primary litigation.}

Of course not every infringement suit turns into a full court battle. That is slight consolation for a firm with limited resources, however, as the costs simply for processing claims is significant. On average, an infringement case which goes to trial carries expected costs in the $3 million range for attorney fees alone (Hunt, 2006).

According to Bessen & Meurer (p. 121-123, 2008) there are two kinds of infringement
suits: cheaters who get caught stealing; and inadvertent infringement suits. If a failure to coordinate causes litigation centered around the first group, society has nothing to worry about; the system is functioning well. But if lawsuits contain more of the latter group then patents and their inability to line up are more likely to be a drain on innovation. Unfortunately, only 4% of litigation involves intentional copying (Bessen & Meurer, 2008, p. 126). Consider that litigation of the willful group should be short and rather quick; cheaters are caught infringing more or less red-handed. Then inadvertent litigation will be relatively longer, with longer implying more expensive. Larger firms are more likely to have the know-how, the reach, and the deep pockets to either offensively prosecute unknowing infringers. They are also better positioned to cheat and get away with it; they have the money to survive a protracted defensive position asserting inadvertent infringement. If found guilty, a willful infringer faces treble damages (Lemley, 2008).

Infringement on the part of small firm, whether purposeful or inadvertent, is not a valid strategy. Empirical work though finds that the small firms who are often the owners of foundational patents are sued disproportionately often relative to their larger siblings (Bessen & Meurer, 2008, p. 123). In pharmaceuticals, final innovations are generally not the intellectual children of a large firm, even when it is the large firm which eventually has their name on the product (Scotchmer, 1996, p. 322). Such is the threat that Lemley (2008) finds evidence of companies encouraging their engineers not to read patents. Ignorance serves as a piece of a defense against claims of willfulness. The threat of litigation seems to now push back against the disclosure pillar of the patent system’s justification.\(^{54}\)

Despite the problems, the expansive scope of options for coordination which have evolved speaks to an adaptable system. As noted at the outset of the section, the number of coordination mechanisms is far broader than this research can cover. Augmentations of basic licensing, including compulsory licenses, technology transfers, contingent licenses, cross-
licenses, and patent pools—a type of membership license—are quite developed and well-utilized in the pharmaceutical industry (Grubb, 2005, p.440-457). They cannot all be covered however.55 There are pieces which are immediately noteworthy, and not only on the negative side. Research has uncovered substantial side-benefits, such as the explicit disclosure which often accompanies licensed technology. Significant quantities of tangential information—codified and uncodified—flows with licenses. Uncodified information can come from, for example, any hands-on training included in the license sharing. Training is often necessary so that the licensee can fully exploit and practice the patented technology without reinvesting time to learn (Hope, 2008, p. 174).56

With many potential advantages one might ask why any firm would not license their product in an efficient way? Bessen (2004) shows that when developers hold their R&D costs private there may not be a reason to offer ex ante licensing opportunities. In a more general sense, or where privacy is hard to guarantee, there are other tricks to employ in order to milk larger profits from licensing. These include exploiting submarine patents.

Submarine patents are filed patents which remain unpublished and hence can suddenly emerge, from underneath the intellectual landscape, as valid and enforceable rights even though subsequent innovation may be underway. The surfacing patent suddenly places a competitor in an infringing position although they had taken all steps to avoid that situation.57 Even with a careful survey of the R&D landscape, uncertainty remains after patent

55The literature touches at many points on antitrust concerns, which also lead the current work too far afield. As a concrete example, the inclusion of substitute products in a single patent pool is effectively collusion (Langinier & Moschini, 2002, p. 14). Complementary patents within a pool, on the other hand, are coordination-enhancing. It is not always known what patents will be complementary or substitutable when a pool is formed however. The relation of the inventions to each other could play a deciding role in whether a coordinating agreement is efficient or not (Chang, 1995). While many solutions have been proposed there remains an overriding concern that the pace of business and the unique circumstances of every situation in high technology industries places a high burden on regulation promulgation and traditional enforcement methods (Posner, 2001).

56Petit (p. 78, 1996) notes that these effects sit well with institutional design principles which state that when attempting to promote a benefit (release of information) a complier-centered strategy is more effective than deviant-punishing strategies (stopping withholding).

57Pharmaceuticals do conduct better searches than the average, as measured by the percentage of citations to other patents the applicant providers compared to the examiner (Sampat, 2004). Bessen and Meurer (p. 154, 2008) report Sampat’s findings that 50% of references in an average application come from the
searches. Such is the level of uncertainty, in fact, that patent searches are virtually uninsurable (Bessen & Meurer, 2008, p. 51). Some inventors did use a now defunct patenting strategy to keep unpublished patent applications pending and therefore submerged almost indefinitely on the off-chance they could become an offensive asset (Grubb, 2005, p. 416). With the signing of the TRIPS amendments to the GATT framework under the World Trade Organization in 1995, submarining is far less of a concern. Patent applications in the U.S. are routinely published 18 months after they are submitted. Previously they would not have been published, thus allowing for submarining. With such offensive options, it appears less strange that Lemley (2008) finds firms defensively pursuing a strategy of ignorance of the patent landscape.

Illustrating both a sense of the need to coordinate intellectual property and an ability to do so, a recent survey by the American Association for the Advancement of Science (AAAS) found that among their respondents—nearly a quarter who are involved in fields of research and nearly half of whom had created something themselves they considered eligible for intellectual protection—most had acquired patented technology for use in their own research since 2001 (S. Hansen, 2005). Despite the usability and availability of routes to coordinating intellectual property and R&D, 40% of the respondents to the AAAS survey also reported difficulties in accomplishing the allowances (S. Hansen, 2005). Given that

---


59 Nevertheless, Hope (p. 45, 2008) notes a 2003 survey of the patent rights surrounding a key technology in plant transformation science where the farthest reaching patents could not be determined as the broadest patents still had yet to issue from the USPTO, showing some uncertainty in the ability for leading edge technology to patent.

60 Submarine patents introduced other methods of patent warfare, situated amongst: “clustering” one’s patents around a core patent; “bracketing” a competitor’s core patent with your own patents; following a policy of “blitzkrieg” and rapidly patenting everything remotely related to each other; “blanketing” the process used to manufacture one’s innovations; “flooding” a competitor’s portfolio by stacking incremental variations of your own thereupon; or “fencing” off lines of research (Hope, 2008; Macdonald, 2004).

61 Of potential note to the pharmaceutically-inclined reader, the most used device for transferring intellectual property in the biosciences was the material transfer agreement (MTA) while exclusive licensing was the least used method (S. Hansen, 2005).
under patent law in the U.S. allows all persons named on the patent to have full rights over the patent, each of whom can therefore grant licenses independent of the others named (Grubb, 2005, p. 397). And above the sheer numbers problem, the literature states that most problematic clause to negotiate is the field-of-use clause (Hope, 2008, p. 49), evidence of access is remarkable.

There remains, of course, difficulties. For example, even where licenses are achieved the field-of-use clauses, which place conditions on what can and cannot be done with the license, can create speed bumps within R&D conducted under the ostensible success of achieving the license. Today there is a human and time cost to bridging gaps and a need to consult with licensing professionals (Hope, 2008, p. 47,50). Green & Scotchmer’s (1995) modeling naturally left out the complexity and transactions costs of licensing for all sectors, let alone pharmaceuticals. But their model illustrates the ability for improvements to capture property rights claimed by previous inventors and how it acts as a sort of “compulsory license.” And there is an upper bound on coordination, especially within a single industry, where a degree of intellectual coordination becomes anticompetitive.

Here again we have identified a dimension of interaction which patents must somehow balance. An emerging irony is that patents seemingly create two problems for every problem they solve. The reality of granting property-like protection to ideas which resist property-like description of their boundaries lands the simple balancing, introduced in the patent narrative, in troubling outcomes. Coordination of innovation becomes necessary as do

---

62 35 U.S.C 261, 262
63 Hope (2008) details that the process of licensing often involves more than one set of property rights and trade secrets, and extends across several rounds, including “both formal and informal instruments [...] confidentiality or nondisclosure agreements, material transfer agreements, option agreements, term sheets or memoranda of understanding, and, increasingly, agreements to negotiate (Hope, 2008, p. 47,220).”
64 There is a sub-literature on compulsory licensing. The U.S. though is quite hostile to the idea in general, although it is allowed in the drug context. As far back as the Vienna Congress of 1873 the United States opposed the adoption of a compulsory licensing clause in international patent law allowing for compulsory licensing of patented inventions when the public interest required it (Penrose, 1951, p. 47). Prior to World War II, the UK allowed for virtually any patent on medicine to be licensed on demand (Grubb, 2005, p. 19).
65 For an introductory discussion of all issues surrounding patents and antitrust concerns, see Kaplow (1984) and recently Scherer (p. 33-36, 2007b).
means for directing the remunerative material to respective coffers.

Nevertheless, all additional dynamics have proved amendable to the sort of analysis which set out to optimize the simple narrative. The topics introduced in this chapter and addressed by the economic literature simply point toward a more complex patent narrative. The complete narrative would incorporate the coordination of patents, the disincentives of over-segmentation of property, and the cost of legal intervention if the goal is a truly optimal innovation policy. Though that is too broad for this research, the discussion serves to highlight the incentives experienced by firms in various circumstances. In order to narrow down the set of various circumstances, the next chapter turns towards describing the pharmaceutical industry specifically. There, it is the underlying science which by and large creates the industrial structure. It also influences the regulatory structure for drugs addressed in later chapters.
Chapter 3

The Pharmaceutical Industry

...redundancy and multiple use are the handmaidens of creativity

Stephen Jay Gould

This chapter surveys the structure of pharmaceutical industry as it pertains to the regulatory structure that governs the industry's outputs. The most important structural note is the recent division of outputs between traditional pharmaceuticals—small-molecule drugs—and and the new biologic medicines—large molecule proteins. The latter are often term biopharmaceuticals.¹ We start the discussion with traditional drugs because they remain the more familiar category and, moreover, were the products of the science that supported the Hatch-Waxman Act of 1984.

The fact that the major structural division follows a scientific division necessitates a digression into biology. The scientific differences between organic chemistry, which forms the backbone of traditional drug R&D, and genetics-based biotechnology, which founded the

¹Although not discussed here, the medical device sector is also a part of the pharmaceutical industry, when viewed as providing therapeutic innovations above the general support tools of the medical profession. Their innovations have their own regulatory pathway, though there are significant differences both in terms of regulation and underlying innovation incentives. They would be best be analyzed in a similar but separate fashion.
biologics industry, helps to explain both the industrial character and subsequent regulatory issues. The scientific division is most important to the latter regulatory discussion because, as we will see, it has not created a true bifurcation in the drug industry.\footnote{Pisano (2002)}

The current view of the traditional pharmaceutical sector finds a mature sector coping with the expiration of the bulk of their patents. They are simultaneously struggling with the limited developments from their R&D pipelines. The annual number of approved products declined from 70 in 1997 to 20 in 2005 (Branna, 2007, p.74). The traditional side's overlap with emerging- and currently-approved biologic therapies, notably still under patent protection, is expected to support some of their revenue. Financial issues aside, the expiration of patents has certainly not diminished the sector's reliance on intellectual property as the guarantor of revenue.\footnote{As an illustration, Bessen & Meurer (2008) routinely separate the pharmaceutical industry from other R&D sectors in their general empirical work.} Their continued reliance on the patent system has led to significant innovation in the application of existing patent law and the exertion of control over parallel branches of law. A penchant for pushing profits into legal innovation—or advertising—rather than scientific innovation has met with no shortage of popular derision.\footnote{I.a., Hope (2008,2004); Bessen & Meurer (2008); and Jaffe & Lerner (2006).}

The downswing in the mood surrounding pharmaceuticals is not however, a new problem for Big Pharma. The reality of developing "miracle drugs" fizzled before. The modern pharmaceutical industry took shape in the 1950s under the unbridled optimism of the sulfanilamide (sulfa drugs) and antibiotic revolutions. The encompassing success of the early century's attack on bacterial scourges and the hope the pioneering techniques brought to challenging other diseases cooled quickly into the day to day toil of research.\footnote{For a history of this important time, see Lesch (2006). In fact, the cooling of sentiments regarding the potential of bacteriology and chemotherapy descended to downright gloom from the mid 1920s to the mid 1930s, just before the sulfa drugs entered the world and changed medicine forever (Lesch, 2006, p. 35).} It was in
this period of realism though that most of the science which currently transforms modern lives came into being. It also brought the intellectual property dynamics and innovations which currently play such a supportive role. More importantly, the march slowly brought the industry into biotechnology.

How far off the next wave of miracle drugs remains however a matter of speculation.\textsuperscript{6} The R&D process itself is still a black box for researchers as are the processes which drive information into the product pipeline (Gassmann, Reepmeyer, & Zedtwitz, 2004, p. 65). What we do know, however, is that intellectual property affects what goes into that black box and we can observe what comes out. Understanding the industry as the confluence of the science going into the R&D box, the law which gives control over the intellectual innovations int that R&D, and regulation as to how and which products make it to the market to generate real remunerative profits for the up-front investments therefore steers this chapter to the science side of the industry. Bookended by the exploration of the patent system (Chapter 2) and the following exposition of the regulatory structure (Chapter 4), a history and contemporary view of the pharmaceutical sector completes the perspective required for economic modeling and policy optimization.

\section{3.1 An Overview of Traditional Pharmaceuticals}

The pharmaceutical sector is truly remarkable in size and scope. The patent landscape has built for the surviving pharmaceutical developers a magnitude of corporate wealth estimated to be up to 33\% of the total value of all corporate assets across all sectors—an astonishing $5 trillion dollars (R. J. Shapiro & Hassertt, 2005). Pharmaceutical patent portfolios constitute a large share of that figure (Bessen & Meurer, 2008, p. 113).

\textsuperscript{6}Predictions from scientists involved in exciting work still generate near daily attention-grabbing headlines. For example, recent claims of Harvard-researcher and Sirtris Pharmaceuticals co-founder David Sinclair, whose early research sparked headlines and a boom for red wine and the substance—resveratrol—contained in red grapes. In large enough doses, the substance promoted longevity in mice (D. Ho, 2008). Understanding how and why has however so far not led to more than a boom for the vitamin industry and red wine consumption.
The shape of the pharmaceutical industry therefore reflects the developments in science which has affected human understanding of chemical interactions with the human body. The industry is young although many of the names which figure prominently into its history go back to the beginning. The foundational firms which established themselves from 1850 until 1945 were not, however, in the business of developing chemicals for use in humans. They were instead focused on the commercialization and mass production of industrial chemicals (Burhop, 2008; V. A. Gilsing, 2005; Santos, 2003).

Although they became pharmaceutical giants, the chemical giants did not set out aiming for an understanding of pharmacological mechanisms (V. A. Gilsing, 2005, p. 7). The tools of organic chemistry established the industrial side to chemistry first. The chemicals and “dye-stuffs” industry emerged with only the capacity to create and manufacture chemicals on a previously unknown scale. The industry had, up until the early 20th century, focused on dyes, photographic processes, and coal tar chemistry (Lesch, 2006). It is perhaps serendipitous that the dyestuff industry, born in 1856 with the first artificial dye—mauve (aniline)—had not already split from a chemical industry. It is after all the use of dyes to stain microscope slides where science realized that dyes had metabolic effects on the microscopic cells themselves (Lesch, 2006, p. 15).

The pharmaceutical seed germinated, for better or for worse, during the World Wars. Whether there are causal threads or pure serendipity, the coincidence of the realization that chemicals have beneficial properties when used in humans happened as society plunged through a time of needing these beneficial therapies. It was the need that drove acceptance of the paradigm. Lesch’s (p. 200-203, 2006) history of the sulfa drugs is indispensable in understanding the ignition of the industry at this time and moreover informs the modern reader of the strangeness of the idea of ingesting a chemical to heal. The term chemotherapy retains a sinister quality in modern usage and it was equally alarming to the pre-war population when it was introduced as general terminology for all the application of pharmaceuticals (Galdston, 1943).
The modern pharmaceutical paradigm emerges with penicillin at the end of World War II (Malerba & Orsenigo, 2002, p. 669). Penicillin followed on the success and subsequent investment of the sulphanilamide (sulfa) class of drugs which appeared in the late 1930s (Lesch, 2006). As pharmaceutical endeavors began, the availability of patents for intellectual property protection made practical a shift from the bulk-preparations of the chemical industry, where the large integrated establishments package, sell, and advertise their own preparations (Temin, 1979, p. 442).

Though the transition from chemical to pharmaceutical industry happened rapidly, spurred by the late war efforts, it should not be forgotten that the first era of the pharmaceutical industry embodied very little of what modernity might recognize as pharmaceutical development. Without the work done to build and manage the infrastructure of chemicals manufacturing, the sulfa drugs’ appearance at the end of the World War period would not have been at the scale required by the war efforts. Even though its effect, or the small influx of penicillin during the end of the period, was not overwhelming for the war effort, the exposure of the armed services to “chemotherapy” certainly spurred the acceptance on the home front as the industry moved into its second epoch.\footnote{See Cavers (1939) for a contemporaneous discussion.}

The transition from chemicals to pharmaceuticals also marked the end of the individual inventor and laboratory tinkerers (Dutfield, 2003, p. 7). Tasks in the laboratory were then divided among specialists and technical staff (Grubb, 2005, p. 382). Chemists spread their focus between the identification of novel chemicals, the industrial-scale synthesis of usable entities, and the development of hunches as to what chemicals interact with the body (V. Gilsing & Nooteboom, 2004; Pisano, 2002). The “heroic theory of invention,” where single inventors bring forth their creations (Diamond, 1997, p. 244) was however an underlying theme of the original patent narrative. Although risk and uncertainty remained even with the emergence of in-house, managed R&D, there was now a pooling of risk as well as of rewards. The era found chemical development programs vertically integrating within
the firms we know today (Temin, 1979). However, with a lingering labor theory of value, the changes did not create any concerted challenge to the patent system (Mossoff, 2001, p. 1274).

The sea change in political economy during the the Enlightenment had only shifted the moorings of the system to new pylons (Chapter 1). The new supports were as amenable to a firm-as-inventor as a personable-garage-tinkerer\(^8\) so long as the government's power to grant a monopoly was checked behind the USPTO approval process. In light of the lack of changes during the Industrial Revolution, the lack of unique attention in the chemical-to-pharmaceutical transition is unsurprising. It stands to argument that, with the age of machines slowly winding down into the electrical age, the precursor of our current age of information, it was the success of the drug industry which kept the momentum of the patent system.

Patents were, after all, a small concern relative to the emerging issue of safety of drugs. Up to the Federal Food, Drug, and Cosmetic Act of 1938 (FDCA), any non-narcotic drug could be purchased without a prescription (Temin, 1979). Pharmaceutical firms had no direct incentives to interact with doctors or the nascent health system. This was a period of strong involvement by pharmacists, who, broadly speaking, compounded the bulk provisions for consumer benefit. The FDCA was precipitated by the thalidomide disaster, where a sedative was found to cause extreme birth defects when used by expecting mothers. The FDCA mandated safety tests be conducted to prevent such tragic misuse. The modern perspective on oversight took shape thirty years later with the Kefauver-Harris Drug Amendments [to the FDCA] of 1962. The amendments added a mandate to test effectiveness as well as safety so that a drugs availability in the pharmacy and to prescribers carried an expected outcome. The 1962 legislation also established the form of the FDA we see today (Desrosiers, 1990; Mossinghoff, 1999; Fox & Bennett, 1987).

The development programs of the post-war years were relatively more formal than the

---

\(^8\)Perhaps garage inventor is an anachronism which should be shifted to barn-inventor for the period?
chemical era but they remained reliant on random screening of molecules for therapeutic activity. Random screening involves exposing naturally and synthesized chemicals to test tube and animal experiments (Malerba & Orsenigo, 2002, p. 669). The random nature led to stories of discovery of therapeutic classes while searching for solutions to entirely different problems (V. A. Gilsing, 2005, p. 7). But the organic chemistry paradigm remained the backbone of research during the early stages of the pharmaceutical industry proper (Malerba & Orsenigo, 2002, p. 669).

Interestingly, the growth of firms at the time did not reflect increases in profits relative to their contemporaries. Early growth rather came from the shifting of profits from small firms to the large ones (Temin, 1979). There are parallels with modern mergers (LaMattina, 2011; Granier & Trinquard, 2010). The fact is relevant because today’s mergers are riding the changes from another scientific evolution analogous to the chemistry-to-pharma change at mid-century. Although much work in pharmaceutical R&D is directed, pure research in pharmacology exists with an aim to remove R&D uncertainty. The holy grail is a “surgical” knowledge of molecular interactions within the body. The ever-present incentive to remove the randomness from random screening payed dividends at the end of the century in the emergence of biotechnology.

Biotechnology is understanding that allows “the attributes of cells, such as their manufacturing capabilities, [to] put biological molecules, such as DNA and proteins, to work […]” (BIO, 2006). Although biotechnological understanding enters the biological sciences with genetics, biotechnological tools only trickled into the the professional pharmaceutical repertoire from the 1970s onward. It did not succeed in creating a new plateau of pharmaceutical endeavors until the the turn of the millennium (Malerba & Orsenigo, 2002). The application of genetic understanding had to wait for the advent of computers which could handle the quantity of information. Even with the machines, Henderson et al. (1999) note that the transfer to computers and, specifically, automated screening, did not lead immediately to the accuracy and cost-savings predicted by the level of understanding granted
by genetics.\textsuperscript{9} Nevertheless, new methods generated tremendous amounts of raw biological data (Gassmann, Reepmeyer, & Zedtwitz, 2008, p. 60).

The data explosion itself drove growth in jobs among workers in “knowledge production”—in pharmaceuticals and across the U.S. economy of the time (Galbraith, 2012)—and thereby catalyzed another industrial shift through changes in science. It is estimated that over 5 million people actively work in knowledge production, a number equivalent to 90\% of all scientists who have ever lived (Gassmann et al., 2008, p. 60). Industrial chemical production had not only scaled their output, professionalized the management of R&D, responded to changes in regulatory environments by managing clinical safety trials, and finally changed their understanding of the foundations of biology, but they now caused an explosion in the generation of raw information. Of more specific import to this research, the information generation has led to significant increases in patenting activity, activity which takes place parallel to the patenting of molecules proper.

The biotechnology shift did not split the sector into pharmaceuticals and biopharmaceuticals, however. The development projects of pre-biotechnology efforts, which focused on small-molecule therapies—or Small Molecule Drugs (SMDs) as they are often referred—fall into the regulatory category of “new chemical entities” (NCEs). These are the medicines common to current consumer usage, those whose scientific roots remain in the organic chemistry paradigm. Biotechnological pursuits have created an initial wave of “biologic” medicines, which are generally large molecules—proteins—which for historical reasons enter the FDA regulatory scheme under a different route. The biotechnological development though remains largely side-by-side in the pharmaceutical sector.

We explain the specifics behind relative size differences later. For now, this short introduction paints both a adequate current picture of the pharmaceutical industry and lends an understanding of the current inertia and direction of evolution.\textsuperscript{10} In order to link this firmly

\textsuperscript{9}See also Gilsing (p. 9, 2006).

\textsuperscript{10}For a detailed history of the evolution of the pharmaceuticals industry along the three phases of chemical, pharmaceutical, and biotechnological, see Galambos & Sturchio (1998, 1996).
to the Chapter 2, the next section views industry evolution through the lens of intellectual property.

3.1.1 The Evolution of Big Pharma

The wonder drugs to emerge around World War II were not patentable due to prior art and their existence in nature (Temin, 1979); they were either existing ideas re-applied or discoveries of natural content not refined by the hand of man. The patenting of streptomycin—an antibiotic—in 1948 sparked the patenting engine. With intellectual property control came the management of R&D decisions as expectations of monopoly gains channelled the research departments' focus toward patentable outcomes.\textsuperscript{11} The success of the combination of intellectual property and industrialized pharmaceutical R&D now delivers fifty billion dollars in revenue (Dudzinski, 2005, p. 156).

The financial weight does not however entirely differentiate between success in managing drug development and success in making drugs (Feki, 2005). Feki (2005) notes that there has been loss of scientists and doctors as the CEOs and board members of drug companies. They were largely replaced by business executives.\textsuperscript{12}

Unfortunately, recent high-profile market failures stemming from safety concerns,\textsuperscript{13} and many less-publicized, but nevertheless expensive, late-stage clinical trial failures indicate that managing R&D has limitations (Feki, 2005, p. 3-4). By 2000, analysts found 80% of drug projects failing in clinical trials (A. Rai, 2001, p. 190).\textsuperscript{14} There is evidence of significant differences in success rates across therapeutic categories (Pammolli, Riccaboni, & Magazzini, 2007; Danzon, Nicholson, & Pereira, 2005), raising the potential that scientific understanding remains homogeneous across R&D projects to an extent that limits effective

\textsuperscript{11}Eisenberg (2003) elaborates on the ways in which law and regulation direct R&D.

\textsuperscript{12}Sometimes referred to in popular literature as a "lab coats to suit coats" transition.

\textsuperscript{13}Notably, Merck's catastrophic withdrawal of Vioxx and competitors' subsequent removal of other COX-2-based pain management therapies.

\textsuperscript{14}Some more recent statistics support a trend: 9,750 of every 10,000 molecules screened in basic laboratory research are discarded as potential therapeutics, only 10 of the remaining 250 enter clinical trials, and only 1 of those 10 succeed in getting FDA approval (Feki, 2005, p. 5-6).
decision making.

In popular press, there has been a slide into a pessimistic mood, a mood captured under the loss of the blockbuster model of pharmaceutical management. The blockbuster model pushed R&D efforts toward large markets with broadly efficacious therapies. The largest pharmaceutical market—the US—also steered efforts towards the large markets populated by well-insured patients. The mega-revenue generated by the blockbuster successes, defined as a drug with annual sales exceeding $1 billion, to smooth over the failures in R&D and the less-well received therapies composing the firm’s product portfolio. With R&D expenditures topping $50 billion annually, profits of that magnitude were seen as necessary to sustain the industry.

While sports coaches have long known the dangers of “riding star players,” the lack of new blockbusters does indicate that pharmaceutical firms failed to a degree to develop enough recruits—or recruitment programs—while their stars took the field. Though hyperbolic in omitting the uncertainty of development, the hopes of biotechnology kept prospects optimistic for some time. Unfortunately, many of the biologic candidates became expensive late-stage failures rather than promising rookies (Hope, 2008, p. 254).\textsuperscript{15} The growing pessimism has not however hurt the legal side of pharmaceutical management.

Without new drug approvals, firms must either increase the profit flows from existing drugs or extend the existing profit profiles past their current patent expiration dates. There has been some successes on both fronts. The successes of legal departments in extending IP were shown in Chapter 2. On the other front, the sector has successfully re-engineered some older products that already have FDA approval (G. Walsh, 2006, p. 770), resulting in increases in the size of the market for the remainder of the patent period. As we discuss in Chapter 4, if the re-engineering required additional clinical trials prior to sales, that time may come back to the patent owner, extending the effective date of expiration of market power.

\textsuperscript{15}See also DiMasi & Grabowski (2007).
Perhaps most notably, Big Pharma also transformed consumers' inactive role in prescription usage through direct to consumer advertising, where legal (Feki, 2005, p. 14). In 2001, expenditures for consumer advertising were reported at $2.7 billion and promotions directly to doctors were $5 billion (Scherer, 2004, p. 928). The advertising push is part of the reason that sales of branded drugs, or drugs which are the first of their kind on the market, retain the lion's share of drug profits (Bhattacharya & Vogt, 2003). Generic drugs, which enter the market after patent expiration on branded, innovator drugs, are chemical copies of the of the original. The expiration of an innovator's patent is seen as the end of market power and the entrance of generic copies theoretically brings the competitive pressures desired to expand consumer surplus.

Generics may make up the majority of prescriptions filled but are only a small portion of total sales (Kotlikoff, 2008, p. 12). The large numbers reside with the branded drugs, with direct-to-consumer advertising and "detail men" who bring targeted persuasion to prescribers. The outcome was the blockbuster model, and generated marketing expenditure inertia, ballooning from trivial expenditure after World War II to roughly 15% of sales receipts by 1972 (Temin, 1979, p. 431, n. 4). Dutfield (p. 242, 2003) reports that from 1995 to 2000, marketing staffs in pharmaceuticals grew from 55,348 employees to 87,810 while research staff stagnated and dropped slightly, declining from 49,409 to 48,527. Though often discussed pejoratively, especially with regard to the new direct-to-consumer portion of advertising, advertising for doctors is not new. We do not discuss the nuances here however, but research shows that the U.S. is in the average range as far as spending on advertising to doctors (Berndt, Danzon, & Kruse, 2007).

In summary, the horizon is not yet bleak for the sector. In a recent study by DiMasi, et

---

16It is legal in the United States and New Zealand.
17They are not always identical. See Chapter 4.
18There are higher profit margins in the U.S., however. Also notable, prior to the 1938 Food Drug and Cosmetic Act most print advertising appeared in popular press. Ads switched to technical medical literature after enactment (Temin, 1979). The side of advertising directed at prescribers is sometimes termed "detailing" (Scherer, 2004, p. 928).
al. (2003), the recent slower rise in the number of preclinical and clinical testing phases of research—small-scale trials in non-human subjects and small groups of humans—relative to late-stage clinical trials with large human populations, could be explained by appealing to better discovery techniques. That might indicate more robust successes soon. The rise of in-house bioinformatics departments, dedicated facilities for managing and correlating protein sequencing data, also lends evidence in this direction, by signalling better data management and data mining supporting the projects which do emerge into later development stages (Hope, 2008, p. 255). 19

The optimistic view is also bolstered by reference to the competition in R&D currently fomented by the remaining industrial structure. The multinational firms who usually fall under the umbrella term “Big Pharma” are the remaining laboratories who had the skill—and good fortune—to ride the blockbuster drug profit wave and remain the top name on a long string of mergers and acquisitions. Nomenclature aside, the visually top-heavy industry is notably un-concentrated in competition (Gassmann et al., 2008, p. 11). 20 As of 2001 the top 10 of Big Pharma still had less than 50% of the total market share (Gassmann et al., 2004). Concentration emerges within therapeutic agent segments however. Any successful product faces only a handful of the pharmaceutical firms for control of their therapeutic area. Therefore, traditional concentration ratios remain low while the concentration ratios within therapeutic markets can be much higher (Malerba & Orsenigo, 2002).

Further dilution of traditional concentration measures comes through the churning sea of generic manufacturers, biotech startups, and independent laboratories which surround the core. Generic manufacturers, many of whom came to the game through the Hatch-Waxman Act’s provisions which lowered follow-on entry post patent expiration, continue

19 Friedman (p. 53, 2008) also notes that, while in-house and specialized bioinformatics firms may play a role in industry dynamics via injections of more and more information, their information and computer programs are more difficult to protect, especially via patents, than drugs and research platforms. Thus, one should be careful when incorporating informatics discussions within discussions of the industry’s relationship with patents and IPR.
20 Also Santerre and Neun (p. 435-436, 2004); Feki (2005); and Gambardella (2000).
to develop production advantages and sales strategies to compete with the first-mover and perceived quality advantages of branded products. To focus on the PTR provisions of the IIWA, however, we limit most of our discussion to the innovator firms.

Small laboratories, of which the new biotech firms are members, may only have one or two pharmaceutical candidates in development—preclinical or clinical—at any time. Their entire asset pool may consist only of the intellectual property rights around those candidates (Hope, 2008, p. 271). University start-ups, cumulative mergers, spinoffs, and NRDOs (No Research, Development Only) firms can all be part of the small-firm pasture around the Big Pharma city. The fine-grained fragmentation around Big Pharma also makes the core appear more coherent in relief (Grubb, 2005, p. 409).

Currently, this fringe thrives by providing comparative advantages at lower scales while Big Pharma leverages their capabilities in both development and the administration of clinical trials (V. A. Gilsing, 2005, p. 9). Start-ups will necessarily run into the regulatory hurdle. Taken together with the present drought of internal R&D, one can see the impetus for mergers. Flush with cash from the blockbuster model, start-ups reaching a proof-of-product stage have an enticing product for their larger siblings. Knowing that they will soon run into a regulatory wall where their agile skill set in research will have no impact, there is a significant incentive to sell their primary intellectual property. After the first wave of pharmaceutical consolidation in the 1990s, there was a second-wave of large mergers. Following the first biotech successes, established firms who did not develop external relationships with startup biotechs lost ground (V. A. Gilsing, 2005; Gambardella, Orsenigo, & Pammolli, 2000). That set up even larger mergers between the previous giants (LaMattina, 2011). The sea of agile start-ups and university spin-offs, however, remains as

---

21 One might also say they are all surrounded—and supported—by another level: the university research infrastructure (Roijakkers & Hagedoorn, 2006). Much of this structure owes to the Bayh-Dole Act of 1980, which allowed Universities to take patent ownership of research performed under government grants. See Mowery et al. (2001).

22 The characteristic has led to the term Dedicated Biotechnology Firms (DBFs) (V. A. Gilsing, 2005, n. 9)

23 See, for example, Friedman (p. 17, 2008).
the core contracted.

From this perspective, biotechnology was not only a scientific revolution but an organizational revolution. The tools opened up the pharmaceutical development process to small entrants. This may be limited up to the point of beginning the expensive human trials regulatory climb, but the spin-off development projects currently heading R&D pipelines brought in-house R&D outside again. The incumbents and small firms appear to form a symbiotic relationship (Malerba & Orsenigo, 2002, p. 698).24

Biotechnology also expanded the research horizons. The competitive fringe is notable for the sheer range of research projects (Filson & Masia, 2007, p. 330). Although the successful startups have not, by and large, emerged as a new generation of mid-level or dominant firms within a therapeutic class, they are poised to play a strong role in both basic research and later-stage product development relative to their predecessors (Gassmann et al., 2008, p. 53).

The coexistence of firms of many sizes remains a departure from the old and a move toward a more heterogeneous industrial landscape (Hope, 2008, p. 246). The heterogeneity may remain bounded in the pre-clinical research and development area, but to the extent that work done prior to human trials streamlines the human trials and, most importantly, minimizes failures from unforeseen adverse reactions which emerge statistically as the size of the test pool grows, the early-stage competition should yield more positive and varied therapeutic outcomes than a concentrated in-house model. This all stems from biotechnology and our increasing understanding of genetics.25

This section illustrates the result on industrial structure that the underlying science

---

24 In simulations of industry evolution, even extremely hostile technologic conditions proposed to give the agile entrant biologic firms a competitive edge failed to lead to situations where the large incumbents were unseated from their leadership position (Malerba & Orsenigo, 2002, p. 699)

25 One could credit this more specifically to pharmacogenomics. Pharmacogenomics is the emerging science of applying genetic research to the science of drug development, a technique promising the science of drug therapy and even individualized drugs (Phillips, Veenstra, Oren, Lee, & Sadee, 2001; Murphy, 2000; Rimmer, 2008). The terms pharmaco-genetics and pharmaco-genomics refer technically to two different phenomenon. The former is the study of correlations of a single drug with multiple genomes, and the latter of multiple drugs with single genomes (Y. Friedman, 2008, p. 57). See also Kalow (2005).
catalyzed. The next section turns to discuss the science in more detail. The goal is ground
the popular narrative for the economic reader because it plays a significant role in the
regulatory discussions to come. The scientific introduction reveals the features which are
most pertinent to the regulatory interactions and, in turn, steer the continuing evolution of
industrial and pharmaceutical firm structure.

3.2 Biotechnology

Biotechnology may be a relatively new industry but there are many products of biotechno-
logical origin of which consumers are already aware. The most common are vaccines. The
original methods for producing antitoxins and antibodies utilized living beings—human
or animal—as a manufacturing process. The product was then extracted from the host’s
blood. This “serum therapy” lost prominence with the emergence of the sulfa drugs in
the 1930s and then with the success of penicillin and other small-molecule antibiotics as
research shifted toward organic chemistry (Lesch, 2006, p. 273).

Our current, deeper, understanding of biologic methods of action are coming full circle,
now tuning the original serum production ideas to finer and finer precision (Dudzinski, 2005,
p. 190).\footnote{Notably, the serum-based biologics identified with older vaccines fall under what Dudzinski (p. 186-7, 2005) terms “biologic biologics,” to separate them from “biologic drugs.” The current research focuses on the
latter category. Biologic drugs, are formally defined as “biological macromolecules, such as polysaccharides, polynucleotides (DNA, RNA) and polypeptides (proteins) (Dudzinski, 2005).” With reference to regulation,
Dudzinski (p. 187, n. 331, 2005) notes that the definitions parallel the classifications already used in the
Orphan Drug Act. 21 C.F.R. §316.3(b)(13)(ii)(C).} Other examples of the host-production model of biologics are beer, wine, and
bread. The machinery employed here is microbial fermentation (BIO, 2006). Keeping this
in mind makes the dividing line between drugs and biologics easier, especially with regard
to their manufacture and the resulting regulation issues.

By definition, the products of the traditional pharmaceutical industry are mainly Small
Molecule Drugs (SMDs). The ”small molecule” part refers “to a discrete chemical entity
that generally would contain no more than fifty nonhydrogen atoms, most commonly carbon,
nitrogen, oxygen, fluorine, chlorine, sulfur, and phosphorus (Dudzinski, 2005, p. 154). Large molecule therapies, in contrast, are the products of the biotechnology sector. To be precise, small molecules are those that have a weight less than 1000 Daltons.\textsuperscript{27} In practice, classic synthetic drugs—broadly speaking, the pills in a typical medicine cabinet—average 300 Da (Schellekens, 2004). Though physical size does not imply simplicity, the active ingredients of most current drugs are small, chemically simple, and homogeneous (Mandel, 2006, p. 19).

Biologic products, on the other hand, are generally much larger, jumping to 4500Da to 270000Da in mass (Schellekens, 2004).\textsuperscript{28} Although there is no bright-line distinction between proteins and small molecule drugs, biologic molecules generally become large enough that they develop secondary (and tertiary) molecular structures. Secondary structures are the outcomes of the way long chains of molecules fold and bend. There are often multiple ways in which a chemically-identical molecular chain can fold and remain stable. On the other hand, modifications of molecules in certain areas of the chain will not necessarily change the folded structure.

The pharmacological problem with large molecule drugs is that the folded, macrostructure is often decisive for the activity of the therapy. An improperly folded protein may be barred from reaching its intended target, much like an ill-fitting key will not open a lock. Once at the target site, the chemical structure of the protein can take center stage, where faults in the links in the chain which did not prevent entry can now cause issues.

Practically, the size of the molecules causes problems in detecting atomic changes in the constitution of the molecule.\textsuperscript{29} In fact, there is a fair degree of heterogeneity even in a naturally occurring sample of a biologic product; biology is not the precise assembly line

\textsuperscript{27}1000 Da, or 1 kiloDalton (1 kDa). A Dalton is unit of mass used to measure small masses, and is equal to the weight of one Hydrogen atom—one unified atomic mass unit (u). A unified atomic mass unit (u) is also referred to as atomic mass unit, and abbreviated amu.

\textsuperscript{28}There are several peptides which are of the size usually covered by SMD size references. See Dudzinski (p. 188-89, 2005).

\textsuperscript{29}“Small molecules do not possess such sophisticated structures harboring several different functions on one molecule. Small molecules like aspirin and acetaminophen can be fully described in terms of their molecular structure. For full identification a limited set of analytical assays [methods which measure concentration or properties of a substance] can be used (Crommelin et al., 2003, p. 4).”
that organic chemistry is manufacturing. Further, science currently has limited tools to visualize the secondary and tertiary structures of proteins. Predicting protein folding is a computationally intense task.\(^3\) In contrast, SMD production can be checked relatively simply for accuracy and consistency, where changes to even one atom in the active ingredient will show up in a spectroscopic analysis and may even affect macroscopic, physical properties (Tucker, Yakatan, & Yakatan, 2008). Bringing the science back into the patent narrative, this all implies an ability to make copies of SMDs which is not shared by Biologics.

The scientific certainty over SMDs is reflected in patent law. SMD compounds may be granted patent protection even if their structure is unknown (Grubb, 2005, p. 220). The law builds on the predictive capabilities of chemistry, allowing the mature scientific understanding to express the substance via organic chemistry principles. Theory leads to a patentable claim on the utility of the molecule or how it is manufacturered.

Biologics, in contrast, are complex enough to defy manufacture by chemical reactions. Like beer fermentation and serum vaccines, biologic medicines are generally manufactured by co-opting living processes within basic cells. The term *recombinant* emerges from this, where existing DNA is *re-combined* with DNA fragments to create a new DNA blueprint which did not exist in nature. Feeding new instructions to a simple bacteria allows the host to produce something different than their normal allotment of proteins.\(^3\)

Many of the first targets for disease treatment via biologics logically emerged from comparing proteins between healthy and sick individuals. The idea was to manufacture a biologic replacement, or a modified analogue, for the missing or improperly-manufactured

\(^3\)For a sense of scale of the problem, a distributed computing program, “folding@home”, was established by Stanford University to harness unused processing power of hundreds of thousands of computers connected to the internet to solve folding problems.

\(^3\)Production also utilizes immune-system cells. Monoclonal antibodies—abbreviated mAb or moAb—allow for the engineering of the proteins commonly known as antibodies. (Dudzinski, 2005, p. 190) The term “monoclonal” refers to the cloning of the original antibody into a perfect copy of itself. The techniques also find application in bio-marking and tracking. The specificity of antibodies works as detectors of equally specific problems. In tracking, engineered antibodies can bind to a very specific substance, allowing researchers to track it through the body (BIO, 2006). We can also bind a toxin to a monoclonal antibody to insure that the toxin only reaches that specific target, keeping the substance inactive throughout the rest of the body.
endogenous proteins (Crommelin et al., 2003, p. 7). Unfortunately for medicine, illness is often the result of a faulty pathway of protein interaction in which the reaction must be newly regulated and not simply augmented by properly constructed molecules. The body’s ability to express any single protein—the “on” switch to our fermentation of biologic products—has several up-regulation and down-regulation paths—the volume switches to the on button. Turning a pathway on, turning the volume up, or turning it down, will all carry secondary effects. In sum, understanding the genetic component of illness is just the start of research; the techniques needed to adjust the malfunction and the manufacturing ability to bring that technique “from the bench to the bedside” still vary widely, defy theoretical prediction, and complicate the management of R&D (Y. Friedman, 2008, p. 60).

3.2.1 Manufacturing

A major hurdle in biotechnology is getting a host cell to do the right thing. While the science of recombining DNA is well-established, coaxing a bacteria cell, like E. coli, to express the desired gene as a protein product is only the first step. The protein often remains within the bacterial cell once it is created, necessitating some disruption of the cell wall and then the separation of the desired proteins from the extra material (Grubb, 2005, p. 254). Admittedly hyperbolic, biologic manufacture has to tear down the brewery to get to the beer.

More precisely, let us assuming the human pharmacological side is known to be safe and effective in order to consider just the manufacturing steps needed. Organelles inside the body’s cells can read a DNA template where certain orders of amino acids—the bytes of genetic programs—instruct the organelle to attach a specific molecules in sequence to create a protein chain. When the cell finishes reading and attaching chemicals in the order prescribed, we have a finished protein chain. A host of factors, chemical and environmental, influence how the chain folds.\footnote{We do not go into details here. Please see Ho & Gilbaldi (2003) for an introduction.}
inserting others, a host cell follows the new instructions and creates a customized protein chain. If other biologic features comply, the chain will also fold correctly for its intended purpose.\textsuperscript{33}

There are situations where a biologic-manufactured protein is excreted by host cells when they are finished with manufacture. Though saving manufacture from tearing down the cellular machinery, other unrequested excrement often tags along (Dudzinski, 2005). Identical DNA segments copied into different host cells—different “families” of the same species of bacteria—can even produce different products (Liang, 2007; Radcliffe, 2004). The output process for most biologics is therefore inherently mixed, creating problems for the transfer to humans.\textsuperscript{34} The problem is multifaceted too, not simply a problem of separating wheat from chaff. We mentioned previously that folding outcomes depend on environmental factors. This implied the environment of the cell manufacturing the protein but it extends to the macro world, where exposure of the protein chain to different laboratory vessels may change the outcome, as well as other macro controls like temperature and forces applied in the medium (Radcliffe, 2004).\textsuperscript{35}

For comparison, SMDs are routinely supplied in purity of 98 to 99\% (Crommelin et al.,

\textsuperscript{33}Note that the recombining of DNA instructions for the host cell to read is different from gene therapy. Biotechnology uses genetic knowledge to create therapeutic proteins, proteins which are then transferred to a patient. In gene therapy, nucleic acids are transferred \textit{directly} to the cells of a patient (Rubanyi, 2001). The goal of gene therapy is to instruct or re-instruct a patient’s own cells with newly encoded instructions. Biologic medicine supplements a patient’s natural compliment of proteins from external sources.

\textsuperscript{34}Additions are sometimes beneficial or part of the manufacturing. This notably includes glycosylation, the process of adding sugar or other molecules to the protein. Many times these are necessary to stabilize the protein or to protect it during its administration to humans. The combination leads to “naturally” heterogenous mixtures relative to small molecule drugs’ production (Schellekens, 2004, p. 406), (Dudzinski, 2005). This is notable to the regulatory discussion with respect to the ability of biosimilar—generic analogues of biologic medicines—to in fact be “similar” to the original.

\textsuperscript{35}Changes in the protein via contact with storage media has also been suggested as a possible cause of negative patient interactions, notably with the Eprex situation (G. Walsh, 2006, p. 774). While not discussed here, there are also problems with delivery, problems which may also touch on storage vessels for proteins. Supposing for the moment that manufacturing is successful delivery pathways for the manufactured protein are usually via parenteral injection or infusion (Raines, 1999). Proteins often become denatured if swallowed or otherwise orally administered (Raines, 1999). The way a protein is changed, degraded, or otherwise modified on its pathway from injection to its endogenous home within the body can have significant consequences for therapeutic outcomes (Crommelin et al., 2003). At times, this works in favor of biologics' safety. Metabolites—byproducts of the the chemical changes drugs undergo in the body—of biological products are often biologically inactive whereas the metabolites of foreign SMDs can be dangerous (Crommelin et al., 2003).
This level of purity is not yet routine in biologics. In fact, even showing the level of purity of a resulting biologic serum is difficult. This science of protein characterization is in need of more advanced techniques if it needs to provide information analogous to SMD production. Bluntly reducing a biologic into its molecular components will not reveal anything of its spatial structure, much the same way as examining a melted house key will not allow anyone to reconstruct the pattern of entry ridges. Currently, “for larger proteins, it is almost impossible to guarantee full equivalence of the protein product characteristics in the clinic, including equivalent efficacy and safety (Crommelin et al., 2003, p. 14).” For the purpose of economic inquiry, we may safely assume that the progression from success on the lab bench to full-scale industrial production is not trivial for either innovators of for generic followers (Y. Friedman, 2008).

Reflecting the level of uncertainty, the FDA requires all production facilities to be compliant with current good manufacturing practices (cGMP) and able to show consistency from batch to batch, and from year to year, in their biologic products. While that is shared with SMD manufacturing, the hurdle is steeper on the biologics side. The steepness implies more expenditures in biologics manufacturing.

Base costs for biotechnology manufacturing capacity are significantly higher than SMDs, resting in the range of $250-450 million and requiring three to five years of construction lead time (H. Grabowski, Cockburn, & Long, 2006). This is on top of materials costs that exceed traditional manufacturing by factors of twenty to one hundred times (H. Grabowski et al., 2006). Grabowski et al. (2008) find new manufacturing in biotechnology consuming 3-5 years in physical construction and cost upwards of $250 million. Retrofitting existing production capabilities can still cost $50-$100 million. Grabowski et al. (p. 448,

---

36Protein products like monoclonal antibodies, which by nature entirely specific, raise larger issues (Y. Friedman, 2008, p. 89).

37Recent international evidence from Japan confirmed that “there is a steep [production] learning curve for companies that are not familiar with the development of biologics (Horikawa, Tsubouchi, & Kawakami, 2009, p. 192).” For a discussion on the physical problems of constructing, upgrading, or modifying biologic manufacturing plants, see Molowa (2001) and discussions in Grabowski et al. (2008).
2007) summarize that “[m]anufacturing biologics typically requires the development of specific production facilities and certified processes, a substantial difference from the inorganic chemistry and bulk production techniques of pharmaceutical manufacturing.”

A motivating example comes from Johnson & Johnson’s manufacturing changes when starting a European production line for their erythropoietin product, named respectively *Procrit* or *Eprex*. The example is particularly illuminating of the problems because the company was replicating their own product; the production was merely moving to a new location. The changes were made by the original manufacturer under the full amount of information which led to the drug’s development in the first place.\(^{38}\) Despite that fact, the new prescriptions caused a sudden explosion of complications—pure red-cell aplasia—in patients, a complication not seen in the use of Procrit (G. Walsh, 2006, p. 773-4).\(^{39}\) Although existing characterization techniques could have identified minor differences between one version of the protein and another, that may not have led to information predicting the violent side-effect differences (Dudzinski, 2005, p. 233). The case is often mentioned as the archetype of danger in recent debates over the capabilities of companies hoping to enter markets for expiring biologics with generic versions: if the innovator cannot always replicate their success or catch mistakes, how can the following competition?

There are two aspects of the manufacturing and characterization difficulties which are important to the patent inquiry. First, significant fixed costs to manufacturing changes the competitive entry assumptions usually employed for the period after patent expiration. The bulk production capabilities for generic SMDs entering the market after innovator expirations did not generate significant concern over capital barriers. The second, related, aspect is the assumption of profit erosion for innovators. The patent narrative rests on

---

38 Another example is the handoff of Raptiva between Genentech and Xoma. After Xoma finished clinical tests, the manufacture process was transferred back to Genentech. Despite the open-book transfer and positive matches in analytic and animal tests between the earlier and new production runs, concentrations in human samples proved to be significantly different in the eyes of the FDA, and further studies were required before approval (Liang, 2007).

39 Eprex is manufactured by Ortho Biotech, which is an affiliate of Johnson & Johnson. Procrit is manufactured by Amgen. There is another version of erythropoietin-α which comes from Amgen.
the assumption of insufficient profits in a competitive market to justify up-front R&D investment. Where other barriers exist, profits may exist without patents. While parallel justifications for patents exist, the ability for innovators to continue profiting after expiration certainly impact discussions of PTR. PTR is further impacted by the fact that the difficulties in manufacturing also cause regulatory difficulties, providing a parallel barrier to entry for competitors.

Naturally these are all problems with which science can wrestle. If a protein’s activity depends not just on their chemical composition but on their complicated geometrical structure then we can expect technologies to emerge which can quantify the three-dimensional structure, just as techniques arrived to characterize small molecules. Science has in fact made progress in fully characterizing biologics in terms of “identity, purity, safety, potency, and effectiveness” with analytical tests including high-performance liquid chromatography, mass spectrometry, electrophoresis, nuclear magnetic resonance (NMR), and 2D-Page imaging (Dudzinski, 2005, p. 222-23). They all have their own drawbacks at the moment but can be expected to improve. As such, there remains currently a dependence on in vivo—in an animal—and in vitro—in a glass—laboratory techniques to test biologic outputs (Grubb, 2005, p. 235). (HPLC)

In summary, this section highlights the scientific differences between traditional, SMD development and modern biologic drugs. The scientific differences have already been seen impacting the costs structures and have led to substantial regulatory debates. The patent narrative and any inquiry into the optimality of patents in the sector must take into account the science differences in so far as they change the existence and height of industrial barriers external to intellectual property protection. Similarly, the shortcomings in the science of characterization and the engineering of biologics manufacturing, relative to the mature

---

40See infra p. 17.
41E.g. X-ray diffraction is a powerful visual tool but the specimen under observation must be put into a medium which is far different than that in which it is actually used as a therapeutic. This has implications for the relevance and advancement of in silico testing, where the three-dimensional structure of a protein is fed into a computer for the purposes of predictive modeling.
levels seen in the SMD side of the industry, impact the safety and efficacy mandate of the regulator and thereby change the entry dynamics. Both industrial and regulatory barriers change the expectation of competitive conditions and will therefore impact the tradeoffs experienced by society when incentivizing these markets into existence via patents.

In the modeling of the Hatch-Waxman Act in Chapter 6, PTR is assumed to act on a market where competitive conditions obtain without patent protection. While the assumption is valid for the historical landscape, this section highlights the changes which impact the modeling assumptions while giving direction to subsequent research. To continue guiding directionality, and to introduce the centrality of the regulatory discussion in Chapter 4, the following section introduces generic biologics—biosimilars—as the biotechnological analogue to the generic small-molecule drug. As the HWA changed both the patent incentives of innovators and the entry pathways for generics, biosimilars can be expected to interact in both patenting and regulatory debates toward similar ends.

3.3 Biologics and Biosimilars: The Emerging Situation

The biotechnology industry stands at a juncture where generic versions of their products are beginning to ask for FDA approval. While very few biosimilars—the generic equivalents of biologic medicines—have entered into the U.S market, there is no shortage of academic and legislative discussion.

The rapid spread of the discussion is not without impetus; current estimates find 75% of currently approved biologics ready for biosimilar competition (G. Walsh, 2006, p. 772). Biosimilar versions of many current blockbusters are known to be in development, including insulin, human growth hormone, interferon alfa-2b, and at least ten versions of erythropoietin (Dudzinski, 2005, p. 182-83). Although entrants appear ready, the number of newly approved products—biologic and traditional—has declined from 70 in 1997 to 20 in 2005.

42See also Tucker, Yakatan and Yakatan (2008).
As the entrants which remain are moving toward patent expiration, and as the biosimilars currently have an unclear regulatory path toward entry, legislative concerns which led to 1984’s HWA are coming around again.

It is not an entirely analogous situation, however. The Hatch-Waxman regulations were implemented during a mature phase of the industry, working with a mature scientific base of understanding after decades of organic chemistry development (Hollingshead & Jacoby, 2009). Biosimilars are relatively new and have several relevant problems, of which molecular characterization and manufacturing problems are primary examples. The term _generic_ as used by the HWA and FDA refers to a duplicate drug which is _bioequivalent_ to the innovator’s product. The term does imply that generic drugs are not exact replicas of innovators’ products. They do, however, replicate the results of the innovator’s drug to a near-identical degree. Most importantly, the near-identical outcomes predict and equally replicable safety profile.

Science and technology comes up short in this regard for biosimilar medicines. This generates the name “bio-similar”, clearly distinguishing their relationship with the innovator from their generic SMD siblings. The term bioequivalent means that the drug shows up the same in blood analysis, making it a measure of the steady-state concentration.

---

43There was significant debate on naming issues leading up to recent legislation. Originally most used the term “follow-on protein product”, which was the official FDA terminology, adopted in 2004. See Federal Register 69, no. 50386cf. (Manheim Jr., Granahan, & Dow, 2006). “Biosimilar,” became a shorthand that stayed around. The term “biogenic”, though still used and emphasizes the regulatory connection to innovator’s product and to the previous generation’s “generic” drug terminology, is imprecise.

44As defined by the FDA: “This term describes pharmaceutical equivalent or alternative products that display comparable bioavailability when studied under similar experimental conditions. Section 505 (j)(7)(B) of the Act describes one set of conditions under which a test and reference listed drug shall be considered bioequivalent: the rate and extent of absorption of the test drug do not show a significant difference from the rate and extent of absorption of the reference drug when administered at the same molar dose of the therapeutic ingredient under similar experimental conditions in either a single dose or multiple doses; or the extent of absorption of the test drug does not show a significant difference from the extent of absorption of the reference drug when administered at the same molar dose of the therapeutic ingredient under similar experimental conditions in either a single dose or multiple doses and the difference from the reference drug in the rate of absorption of the drug is intentional, is reflected in its proposed labeling, is not essential to the attainment of effective body drug concentrations on chronic use, and is considered medically insignificant for the drug. Where these above methods are not applicable (e.g., for drug products that are not intended to be absorbed into the bloodstream), other in vivo or in vitro test methods to demonstrate bioequivalence may be appropriate. Bioequivalence may sometimes be demonstrated using an in vitro bioequivalence
and rate of action of a drug (Dudzinski, 2005). As most large-molecule protein products are delivered intravenously they will automatically have 100% bioavailability; there is no intervening metabolism or gastrointestinal absorption to filter the molecule on its way to its therapeutic destination. With SMDs and generics, the intervening biological processes serve as an internal screening test with like-molecules being treated the same.

Naturally, pharmacology can invent new measures here. For example, the European Medicines Agency—EMA, previously EMEA—who began regulating biosimilar entry earlier than the US, looks at a battery of pharmacokinetic studies including the the absorption, distribution, metabolism, and the eventual bodily disposal of the therapeutic, supplemented with pharmacodynamic studies, such as the physiological drug effects. These are supplemental to the bioavailability measurements and are combined to predict the safety profile relative to the innovator (Liang, 2007, p.54). Following from the problems of characterization as well as manufacturing, biologic and biosimilar measurement is not entirely straight forward or predictive.

standard, especially when such an in vitro test has been correlated with human in vivo bioavailability data. In other situations, bioequivalence may sometimes be demonstrated through comparative clinical trials or pharmacodynamic studies, among other methods (Food and Drug Administration (US FDA), 2006; Colucci, Marier, & Ducharme, 2008; Colucci, Pasternyk-Di Marco, Potvin, & Ducharme, 2008).

There are certain classifications of generic drugs within the FDA approval scheme, which tell doctors and consumers, but especially pharmacists, just how identical and/or substitutable, they are. See Dudzinski (2005). In brief, however, drugs are coded in the Orange Book by two letters and perhaps a number, XX#. The first letter is either an A or a B, the former relating therapeutic equivalence and the latter a lack of therapeutic equivalence. The second letter gives specifics. For example, AA tells us that an in vivo test was not necessary to establish equivalence, but rather in vitro tests. AB tells us that a bioequivalence trial is necessary to demonstrate bioequivalence. A number will often be attached to AB results, i.e. AB1, AB2, etc. An “AB1" generic is therapeutically equivalent to one of the pharmaceutical alternatives that serve as reference listed (innovator's) products. Pharmaceutical equivalents have same therapeutic moiety but differ on parameters of form, such as capsule or tablet physical forms (Gelber, 2008). Thus AB1 is therapeutically equivalent to a specific form of the reference listed product, and a test was conducted to demonstrate that. For more detail see Gelber (p. 19-21, , 2008).

The Center for Biologies Evaluation and Research Center (CBER), the body responsible for biologic approval, does not use the term bio equivalence; drugs are considered bioequivalent via bioavailability numbers showing no difference in the rate and extend of absorption of the same dosage. (21 C.F.R. §320.1(e))

See, for discussions, Dudzinski (p. 231, 2005).

The FDA defines bioavailability to mean: “...the rate and extent to which the active ingredient or active moiety is absorbed from a drug product and becomes available at the site of action. For drug products that are not intended to be absorbed into the bloodstream, bioavailability may be assessed by measurements intended to reflect the rate and extent to which the active ingredient or active moiety becomes available at the site of action (Food and Drug Administration (US FDA), 2006)."
This remained until recently the reason to block biosimilar applications in the U.S. When biologic medicines went off-patent, biosimilars could only enter the market through the same hurdles as the original, including performing proprietary clinical trials on the molecule they produced. This is a large difference from the smoothed entry path for traditional generics.\textsuperscript{49} There were some biosimilar entrants but they entered under exception to the rules rather than as trailblazers.\textsuperscript{50} Importantly, these entrants did not receive the FDA’s stamp of approval to be substitutable for the original (Liang, 2007, p. 47).\textsuperscript{51} Without the substitutability rating, the entrants were not assured access to patients already using innovator’s version through the pharmacy substitutions that are often part of—or required by—insurance coverage when available.

Only in 2009 did the\textit{ Biologics Price Competition and Innovation Act} (BPCIA) enter into force as part of the healthcare overhaul which makes up the\textit{ Patient Protection and Affordable Care Act} (PPACA).\textsuperscript{52} Signed into law in 2010, the BPCIA creates a HWA-type tradeoff between innovators, followers, and society. The BPCIA grants exclusivity to innovators and grants shortened approval pathways tailored for biologic uncertainties. However and despite its codification, the BPCIA retains much of the ad hoc nature of biologic regulation prior to the passing of the PPACA.\textsuperscript{53} The clear downside is continued uncertainty for innovators and followers with regard to their prospects of entry and the

\textsuperscript{49}Discussed in more detail in Chapter 4. Proposals were abundant but have always stayed close to the existing Hatch-Waxman pathways established for traditional pharmaceuticals. See Liang (2007) for a discussion.

\textsuperscript{50}Omnitrope, a biosimilar version of human insulin, entered the US and European markets in 2006 and is often cited as the first successful “biogeneric” (Tucker et al., 2008). Omnitrope however earned an FDA “BX” rating, which means that although it is comparable to a reference drug—Pfizer’s Genotropin—it is not therapeutically equivalent.

\textsuperscript{51}Generic drugs either receive an “A” or a “B” rating. A-drugs have been proven bioequivalent. B-drugs may be bioequivalent, but have not been shown to be to degree needed by the FDA, and therefore cannot be recommended as substitutes. The second letter in the rating refers to the method used to establish bioequivalence in testing and the formulation of the therapeutic. For another example, biosimilar productions of Amgen’s epoetin alfa have tried to enter the US market since their patent expiration. Amgen has successfully held off these biosimilar challengers since its main patents expiration through secondary product and process patents (Engelberg, Kesselheim, & Avorn, 2009).

\textsuperscript{52}124 Stat. 119 et seq. For a fuller discussion on the legislation than is attempted here, see Carver (2010).

\textsuperscript{53}The ad hoc structure reflects the one already in place in Europe. The EMA holds itself to a case-by-case approach for each biosimilar applicant (Tucker et al., 2008, p. 60).
prospects for profit.

While a deeper discussion of the new review processes takes us too far afield, this introduction suffices to connect the conversation back to patent law. A inability to copy innovators drug, where the inability is based in science rather than regulatory bars, calls into question some basic tenants of the patent narrative.\footnote{It could also amount to a lack of enablement by the original patentee. “By arguing that equivalence cannot be established for biologics, pioneer industry is stating (apparently without realizing it) that its biologic patents are not fully enabled (Mandel, 2006, p. 4).” See also Corbitt (2008). The Supreme Court, in \textit{Consolidated Elec. Light Co. v. McKeeporst}, (159 U.S. 465, 468 (1895)) noted that “[i]f the description be so vague and uncertain that no one can tell, except by independent experiments, how to construct the patented device, the patent is void.” The sentiment was echoed more recently in \textit{re Vaeck} 947 F.2d 488 (Fed. Cir. 1991): “The first paragraph of 35 U.S.C.S. §112 requires, inter alia, that the specification of a patent enable any person skilled in the art to which it pertains to make and use the claimed invention. Although the statute does not say so, enablement requires that the specification teach those in the art to make and use the invention without undue experimentation. That some experimentation may be required is not fatal; the issue is whether the amount of experimentation required is undue.”} Although disclosure in a patent does not need to be a step-by-step recipe and experimentation, trial-and-error, and reverse-engineering may be necessary to successfully copy the patentee’s original work, a sustained inability to fully realize the invention from the disclosure does preclude the award of a patent.\footnote{For opinions specific to enablement in biotechnology, see \textit{Amgen, Inc. v. Chugai Pharm. Co.}, 927 F.2d 1200 (Fed. Cir. 1991); \textit{In re Wands}, 858 F.2d 731 (Fed. Cir. 1988); \textit{In re Lundak}, 773 F.2d 1216 (Fed. Cir. 1985). Discussed in Mandel, (p. 23, n.148, 2006).}

Moreover and relevant to the current state of affairs where the patent does issue, absent a realistic chance for competition allows the innovator to remain a monopolist past expiration. The patent narrative, as seen from the social balancing act in the optimal patent literature, is impacted directly by the state of the science and the regulations which it engenders. The new regulations are not entirely blind to the situation however.

In awarding market exclusivity, the FDA takes on an expanded role which now includes responsibility for adjusting competition in a therapeutic area. The exclusivity provisions of the BPCIA promise the innovator that the FDA will keep similarly situated therapeutics from competing. Exclusivity operates without reference to the patent status of an original invention; if the entrant treats a certain disease but has no patent, the FDA may now ostensibly prevent non-(patent)-infringing competitors from entering to treat the specific
indication. Although not a new idea, the new durations are poised to take over for patents as the primary guarantor of market power and profits (Eisenberg, 2001).\textsuperscript{56}

The formal introduction of biosimilars regulation, ad hoc though it may be, and the addition of market exclusivity to the package of innovation policy afforded specifically to pharmaceuticals, assures that biotechnology has not as yet finished reshaping the sector. Though the number of impacts to come remains high, and outcomes remain equally hard to foresee, the assumptions which drove the form of the major legislation for the previous generation of SMDs and their generic followers require renewed assessment. In all likelihood, they will require total revisions in pursuit of optimal innovation policy research of the sort accomplished after the HWA.

The revisions dictated by considerations of new industrial organization and changing markets, however, will have to come through the current regulatory process. Our focus therefore turns to the particulars of drug regulation in the United States. Understanding the current situation leads to a more complete picture of where modifications can and will occur. In contrast to the HWA’s enactment, the industry’s R&D budgets are now larger; the market is already heavily populated in many therapeutic areas with traditional drugs and generics; consumers are clamoring for reduced out-of-pocket expenditures; the pipelines of the major drug companies are not yet generating the expected successes; and the U.S. is starting a major revision to their health insurance and health provision infrastructure.\textsuperscript{57} This is the future landscape from which remunerative profits will flow. To continue to provide sufficient ex ante incentives under the still-guiding patent narrative, profits must remunerate R&D expenditures in the range of $802 million.\textsuperscript{58}

\textsuperscript{56}Current SMDs enjoy exclusivity provisions too. They are usually trumped by the remaining and restored patent terms, however and have not been a major focus of research.

\textsuperscript{57}For relevant discussions, see Angell (p. 47, 2005); Rai et al. (2008); Magazzini, Pammolli, & Riccaboni (2007) (esp. Figure 1); and Mervis (2005). Compare also Scherer (2007) who notes that general claims of a drying-up of major pharmaceutical approvals are not entirely accurate and may stem from measurement errors.

\textsuperscript{58}In 2000 USD. The $802 million figure is the most often cited study on the cost of drug development and comes from the Tufts Center for the Study of Drug Development, published by DiMasi et al. (2003). The figure embodies out-of-pocket costs of roughly $400 million and then the capitalization of expenditures on
Biologic markets do seem up to the task, where first-generation treatments can cost up to 22 times more than the average SMD cost, with maximum costs reaching $100,000 USD per year (Kotlikoff, 2008). Costs can be justified, however, with the substantial efficacy benefits over existing therapies (FTC, 2009). Biosimilar versions, should they enter the market as “generic” substitutes can then offer reductions in costs with the same improvements in outcomes. Owing to the difference in science and manufacturing, however, cost reductions are in the range of twenty to thirty percent (Mandel, 2006, p. 6), a notably small reduction relative to that seen between SMDs and generics (H. Grabowski, 2008).

Most importantly, these dynamics all fit into the framework of analysis introduced by the previous chapters’ discussion of the incentives and tradeoffs to patents and which continued here with a view on the industrial organization of the industry. Viewing the industry’s evolution as momentum from the science helps explain the previous phases of growth and lends predictions to the coming changes pressed by the biotechnology revolution. What remains is then to understand the legal and institutional forms which will shape—and be shaped—by the continued flow R&D, discoveries, market successes, and their inevitable improvements.

failed projects. As often happens to overly-cited average figures, the results have seen criticism. Notable problems include the use of self reported data from 10 mostly large, corporations generally acknowledged as part of the Big Pharma core (Scherer, 2004, p. 928). Big Pharma has incentives to inflate their reported costs of R&D (Grubb, 2005, p. 405). Self-reported data has a history of being problematic in the industry (Office of Technology Assistance, 1993). See also Light & Warburton (2005). The range of the figure has been confirmed too, in the average, by Adams (2006) who concluded with a large range of costs from $500 million to $2,000 million
Chapter 4

The Regulatory Landscape

Human beings use the power of scientific knowledge to assert and defend the values and goals they already have.

----------------------
John N. Gray

The Food and Drug Administration (FDA) is tasked with assuring the safety of the nation’s food and medical supplies. While their obligations are manifold, this chapter’s introduction to their duties is bounded to the established approval processes for pharmaceutical products. While we introduce some aspects where change will occur as the biologic pathways become defined in practice rather than just in statute, the focus remains on the established dynamics. More generally, it is the components of the approval policies that shape the incentives for R&D investment which interest us. A sufficiently detailed introduction connecting existing regulation with known incentivized reactions should extend to new additions in a straightforward manner.

The FDA’s regulatory role began in 1902 with the passage of the Biologics Act.\(^1\) Though

\(^1\)Pub. L. No. 57-244. ch. 1378, 32 Stat. 728 (July 1, 1902).
sharing a name with today’s biotechnology, the Biologics Act of 1902 addressed viruses, serums, and toxins of a more organic nature (Lesch, 2006). Prior to this, the FDA’s main role was on the post-market punishment of irresponsible and ineffective pharmaceuticals—snake oils (Dudzinski, 2005, p. 147). The mandate now shifted to pre-market examination of a drug’s safety.

The Biologics Act governed the agency until 1938’s Food, Drug and Cosmetic Act (FDCA). The update brought a significant expansion of the scope of involvement for the agency and included an abbreviated pathway for generic drugs. The early form to generic approval required no human or animal testing for entrants. Entrants merely had to be chemically similar to the branded drug (Desrosiers, 1990, p. 137). The historical note set the stage for the HWA’s revisions nearly 50 years later.

Intervening in the progression though was the Kefauver-Harris Amendments of 1962. They removed the generic pathways as part of a heavy reaction to the thalidomide tragedy. As noted previously, thalidomide, when used in pregnant women to calm nausea, caused severe birth defects. The public outcry was well justified if one takes a literal view of agency’s mandate under the FDCA (Desrosiers, 1990; Moshinghoff, 1999; Fox & Bennett, 1987). Above the blanket removal of a simplified entry protocol for generics, the government mounted a Drug Efficacy Study (DESI) to determine if the drugs already in the market were in fact effective. This efficacy requirement was a significant expansion of the prior focus on safety. It extended backwards to approved drugs and to all new applicants as well. The new regulation would have an impact on the R&D landscape as firms would now have to determine what research showed sufficient efficacy to merit continued research.

The FDCA’s policy for permitting generic versions of pre-1962 drug forms remained part of FDA law. The historical note is relevant mainly in regard to its framing effect on future generic conversations. The existing law served only limited use from 1962 forward because drugs which were grandfathered into the abbreviated pathways had limited- or no-patent protection remaining (Fox & Bennett, 1987, p. 96). It would be two more decades until
generics would have a realistic abbreviated pathway where the did not fact similar costs of entry to the innovators, notably without the benefit of a monopoly market waiting at the end of those high entry costs.

1984’s Hatch-Waxman Act (HWA) finally lowered the barriers for copies of new products once they went off patent, and followed the lead of the FDCA in requiring bioequivalence proof. The HWA did not only give generic firms what they had been missing; the HWA helped Innovators by restoring patent time lost in the long and growing clinical trials times needed for to meet the FDA’s modern safety and efficacy hurdles. While detailing what those hurdles now entail could fill volumes, an economic inquiry operates well focusing only on the explicit costs for the firm and time consumed with respect to their patents, an implicit cost to the time required to complete the trials.

Clinical trials are generally divided into three stages of increasing size, complexity, and cost. In each stage, the FDA asks drugs and biologics to provide greater levels of information (H. Grabowski, 2008). The firm, for their part, receives more detailed information on a potential therapy’s pharmacological profile, which translates into a picture of the prospective market for the drug. Of course, negative indications on safety will quickly damage the expectations for the drug and development may be abandoned. Otherwise, the firm continues to invest in proving the drug’s safety and, just as importantly, defining the drug’s efficacy in the treatment of stated indications. Proving that the therapy works for a wider set of indications is clearly beneficial in the sense of defining a broader market. Even with sufficient safety, a narrow set of indications implies a smaller patient population from which to justify continue R&D costs. Although limiting in a less direct manner than the science itself, regulatory hurdles are no less influential in defining which drugs we see in the market.

These regulatory responsibilities continue to evolve, most notably with the recent Food and Drug Administration Modernization Act (Dudzinski, 2005, p. 157,183) and the Patient Protection and Affordable Care Act. The primary regulatory routes however are well-defined in theory and in practice. With this overview in mind, and especially taking into consider-
ation the historical momentum it lends to the FDA’s institutional evolution, the following sections wades into the four major pathways for small-molecule drug approval. We extend the narrative in the final section to the new and emerging regulatory structure for biologics and biosimilars. Though legal details are omitted, the overview is sufficient to establish the role regulatory involvement plays in incentivizing and guiding R&D projects from benchwork through to the bedside table.

4.1 The Pathways towards Approval

There are four main pathways to FDA approval for drugs. Before getting into each, it is important to clarify what the FDA is actually approving when it signs off on allowing market entry. FDA approval is their approval of a marketing application.\(^2\) The marketing application consists of 1-2 pages of labeling and inserts which must be included with a medication. The clinical trials process, and the pre-clinical work leading up to that, is the task of supporting the claims and instructions made on the labeling of a drug (Fordyce & Cahill, 2002). The claims that a firm seek to have approved on their labeling therefore determine what studies will be conducted by indicating what endpoints must be achieved before the FDA will recognize the claims.

Viewing the approval process as a march to support a labeling claim highlights why clinical trials “begin with the end in mind” (Fordyce & Cahill, 2002, p. 320). If the end is an entirely new therapeutic entering the market to treat illness in an entirely new way, it is easy to see why that project’s approval process will be the most substantial. Where the endpoints are less of a departure from the established, the less substantial the route to approval will be.

Specifically, the most substantial routes are the New Drug Application (NDA)\(^3\) for

---

\(^2\)Marketing Application defined under 21 CFR §312.3(b)
\(^3\)FDCA §505(b)(1), 21 U.S.C. §355
traditional drugs and its sibling, the Biologics License Application (BLA)\textsuperscript{4} for biologic medicines.\textsuperscript{5} Relative to proving the claims of brand-new therapeutic projects, a generic entrant coming into the market via the Hatch-Waxman provisions of the Abbreviated New Drug Application (ANDA)\textsuperscript{6}, or via §505(b)(2) of the FDCA, faces a lower bar. Finally, and with some overlap to the ANDA and §505(b)(2), we have the seldom-used “Paper” NDA route. The entry of the new Biologics Price Competition and Innovation Act (BPCIA) lends new biosimilar pathways too but is only touched on briefly here. All variations however turn on the level of support needed for the medical claims the firm wishes to make about their project.

A policy discussion on regulatory involvement might remain bounded to influencing the industrial landscape if not for its accidental—but arguably unavoidable—trespass into patents terrain. Innovators patent their molecular compounds in the earliest stages of R&D. They update and extend their patent filings as they pass through pre- and clinical-development with the influx of information (H. Grabowski, 2008, p. 479). Pharmaceutical firms are tight-lipped about their work prior to Investigational New Drug Application (IND) filing, which marks the official regulatory transition between preclinical and clinical work. The application itself is required to permit a drug of unproven safety and efficacy to be used in the limited and controlled sense of human trials. With that application comes a release on the tight control of information around the pre-clinical therapeutic endeavor. Although the firm has a long way to go until approval and the beginning of market profits, attempting to start patent protection past this point is unlikely and potentially dangerous.

This all means that the statutory grant of intellectual property protection, which the patent creates for all innovations in all sectors, starts to run down in a unique fashion for pharmaceuticals, culminating in eventual an eroded patent term effective from the date of

\textsuperscript{4}21 C.F.R. §601.4
\textsuperscript{5}Medical devices which have no predicate in the market and therefore must establish a complete safety and efficacy profile similar to NDA/BLA drugs follow the Premarket Approval (PMA) process. This is similarly expensive relative to shortened pathways like the 510(k) application and exemption-based approvals.
\textsuperscript{6}FDCA §505(j)
market entry although begun by, at the latest, the start of clinical trials with the filing of the IND. There is a period of time needed for the FDA to evaluate submitted paperwork too. Compared with the stages of clinical trials leading up to the application, the review process is relatively short. For innovators, the time between application-filing and market entry is refunded in full to the patent term. The distinction between review time and clinical trials time is discussed with the HWA’s provisions below. The analysis however focuses on the time spent in clinical trials and, analogously, on the ability to avoid clinical trials by using the generic approval pathways. We start the discussion with the most demanding of pathways and proceed through the abbreviated distances.

4.1.1 The NDA and Clinical Trials

A New Drug Application (NDA) is the most complete approval pathway.\textsuperscript{7} It is intended to scrutinize new therapeutic products—drugs\textsuperscript{8}—without comparable products already approved for market entry. In the case of medical devices, a class of therapeutics not discussed in depth here, comparables include safety data of similar products to which even innovators may claim reference. For drugs, the difference between an innovator and a follower or generic is the existence of a chemically-comparable substance in the market. As any modification of the substance can be expected to lead to changes, at least in the absence of further information, any change to a chemical compound is assumed to be a different compound without comparisons. Therefore, the NDA is essentially blank application where every question of the FDA is required to be completed by the applicant. The concept has not changed significantly from the procedures followed by I.G. Farben and Bayer in the very first days of the modern pharmaceutical industry (Lesch, 2006, Chap. 2).\textsuperscript{9}

For all intents and purposes, the FDA knows nothing of the chemical and must be

\textsuperscript{7}FDCA §505(b)(1), 21 C.F.R. §314 et seq. Referred to here as the NDA to distinguish from the 505(b)(2) pathway.


\textsuperscript{9}Patenting practices, though, have evolved over the same time period. See esp. Dutfield (2009).
convinced of its safety and efficacy. The convincing requires upwards of 100,000 pages of documentation, spanning all levels of clinical trials (Corbitt, 2008, n. 4).\textsuperscript{11} Filling in every question includes the chemical and manufacturing data, nonclinical and clinical data on pharmacology and toxicology, human pharmacokinetics and bioavailability, statistical data correlating side effects and risk factors, and creating informative labeling.\textsuperscript{11}

For an innovator everything begins with preclinical work. These “non-clinical” tests identify and study links between chemical properties and pharmacological outcomes outside the human body (Crommelin et al., 2003, p. 8). Preclinical work may involve living subjects besides humans and can employ models of human processes ex vivo, exposing tissues and cells the to drug outside of the human body. The progression generally starts small with benchwork models and moves upwards to animal tests. As noted, for the a discussion of patenting incentives, we may mark the time between pre-clinical and clinical studies by the filing of the Investigational New Drug application (IND) with the FDA.

Any chemical that has admirable therapeutic properties must file an Investigational New Drug Application before moving into human studies. The IND application applies to both traditional drugs (NCEs) and biologics (Dudzinski, 2005, p. 184).\textsuperscript{12} In addition to human safety, which is ensured by an appropriate Institutional Review Board (IRB) approval prior to the IND application, the IND is required before the chemical can be moved across state borders. The sale of therapeutics for human consumption is a regulated activity (Duch & Ferris, 2002).\textsuperscript{13}

\textsuperscript{10}The FDA is legally held to only six months (21 C.F.R. §312 (2007)) to wade through the application. In reality it takes on average 30 months (Corbitt, 2008, n. 4).

\textsuperscript{11}21 C.F.R. 314.50, See also Dudzinski (2005) condensing the NDA requirements to six categories: list of ingredients, drug composition, manufacturing specifications, samples, labeling, and clinical data. There is the interesting question of the incentives of pharmaceutical firms to develop their own portfolio of data to support their prospective drug, which in turn determines their financial profits, we note the original model in the United Kingdom of the Therapeutic Trials Committee (Lesch, 2006, p. 133). The TTC would take potential drug candidates submitted by firms and take them further in testing via their own choices in laboratories.

\textsuperscript{12}21 U.S.C. §321(g)(1982); also to antibiotics (Duch & Ferris, 2002). For medical devices, there is an Investigational Device Exemption (IDE) pathway which is analogous to the chemical-therapeutic’s request to begin using the unapproved project in human studies.

\textsuperscript{13}There are even special companies which exist to deal with shipping such drugs and biological samples
The IND process is notably passive; no formal approval is issued. An application may fail establish safety and earn a "clinical hold" however (Duch & Ferris, 2002). Also notable, the IND process is largely unique to the US; international regulatory process do not always include the step (Gassmann et al., 2004, p. 47-48). With the tacit IND approval, the applicant moving into clinical work and must keep the FDA abreast of developments in the studies. The updates occur via annual reports, although much more documentation happens concurrently as part of the trial design itself. Failure to file the formal annual reports causes the IND to lapse; it must remain active up to the new drug application (NDA) filing at the end of trials (Duch & Ferris, 2002).

The IND is the arguably the latest time when the chemical or protein can patented. Initial utility may have been patented early in the preclinical studies but the focus of intellectual property protection becomes more specific as more information becomes available as to the molecule’s therapeutic utility. The most specific and valuable protection should therefore enter into force by the time the firm reveals the particular outcomes of their R&D project to begin down the FDA’s regulatory path. Again, while patents may be filed earlier, the IND juncture serves as a crucial point where secrecy and information move outside the firm’s control (Hope, 2008, p. 56). It is unlikely that a firm would withhold patenting until after an IND application. Concerns about release of information, as well as the requirements of patent law itself, push against any later date.

With an IND in hand—and presumably patent protection on the way—pharmaceuticals proceed to the clinical trial portion of the NDA. Clinical trials last on average from four to six years after the pre-clinical trials. They are the most expensive part of drug development (Hope, 2008, p. 56). Recently, there are new firms emerging which specialize in the process, called “contract research organizations” (CROs). CROs may specialize in specific portions of the clinical development process or may lend expertise to the entire path (Hope, 2008, p. 213). This can include living biological material; “wet-ware” (Hope, 2008, p. 144).
The emergence of CROs, in addition to affecting the sector’s organization, allow for degree of simplification in modeling the development process. We can abstract away from the potential capabilities questions and assume that a firm of any size may contract out for the necessary regulatory experience. In the same vein, interesting chemicals should be able to find a financial pathway to the market through alliances with a large firm if they cannot themselves secure funding in venture markets.

The clinical trial period is generally divided into three sub-phases. Phase I extends the pre-clinical findings of low toxicity in animal studies to humans. Phase I involves only a few—20 to 30—healthy human subjects and pursues a narrow goal of replicating the predicted safety profile. Phase I studies do not necessarily reveal much of the effectiveness of the therapeutic for humans. Phase I is far enough removed from the market that they have only a 10% chance of market approval. That is, a drug’s presence in Phase I is not entirely indicative of an impending market success. The testing here lasts between one and three years, after which the company has enough data to decide whether or not to pursue Phase II tests (Y. Friedman, 2008, p. 142). On a final note, biologics are changing some Phase I structures. For instance, they can be significantly longer than for SMDs (Y. Friedman, 2008). The tests that go into their Phase I studies, however, can be more predictive of later success, or may lend more predictive information as to eventual failures. Thus, extending the first phases of biologic studies can have benefits not available to SMDs (Vernon, Johnson, Hughen, & Trujillo, 2006).

---

15 Some research has noted that the emergence of entirely new firms devoted to managing R&D efficiency and development prospects is a side-benefit of patents (Arora, Fosfuri, & Gambardella, 2001). Akin to the phenomenon that patents foster innovation in things, they might also foster innovation in business forms. See, however, contra Mirowski (p. 142, 2004), who notes that CROs in 2000 captured 60% of research grants from pharmaceutical companies. This is money that otherwise could, in part, be allocated to Universities. The CRO business model can be more efficient and targeted to the business-ends of the firm, whereas the University models had both business and pure-research components. This could prove to be a short-term success for pharmaceutical investment, at the expense of long-term scientific progress.

16 See Lerner, Shane, & Tsai (2003), for discussions on the funding sources or biopharmaceutical R&D, including linkages to the business cycle.

17 Garnering approval from the National Institutes of Health’s (NIH) Recombinant DNA Advisory Committee (RAC), in addition to the FDA can add significant delay to motion.
Phase II clinical trials establish the efficacy of the drug along with identifying and quantifying side effects. The stage takes place in larger human populations, including on average 100-300 participants over two years (Y. Friedman, 2008, p. 143). Here the goal is dose-ranging, searching for safe, tolerable, and effective dose. The larger population allows for adjusting to a placebo control group as well as revealing adverse side effects—major and minor—hidden in smaller populations. Data management requirements rise here, adding costs to each new enrollee and every additional extra day of study (Y. Friedman, 2008).\textsuperscript{18}

As in Phase I, extra efforts may be rewarded by identifying problems before moving into the extremely costly Phase III trials. Identifying and addressing problems is important in two respects. First, abandoning a less successful molecule earlier frees time and resources for other molecules. The problems may be small, but early identification may point the firm toward away from a problematic molecule and toward a closely related molecule. Thus, the identification prevents pursuing a less efficacious molecule into large trials. Most importantly, issues which may prove a hurdle towards approval can be addressed by properly designed Phase II and III trials. A red flag is only a concern if it remains a red flag while the FDA pours over the final application. An indication of a safety problem in Phase II allows the firm to structure Phase III investigations in order to accumulate sufficient data to assuage doubts.\textsuperscript{19}

The decision to push a product from Phase II to Phase III is not made lightly (Roberts Jr., Lynch Jr, & Chabner, 2003). Friedman (p. 143) notes that a study of 200 “encouraging” Phase II cancer therapy results found only 13% proceeding to Phase III. At this stage, the therapeutic has already incurred significant capital costs accumulated from the benchwork through Phase I and II clinical trials over the course of several years. Patent term has eroded for those years, likely having started running since at, the latest, the IND filing. There is now significant scientific data support a decision but it unlikely that the data is one-sided.

\textsuperscript{18}There are also increasing difficulties in locating enough patients with the right characteristics to even conduct a powerful enough study.

\textsuperscript{19}See discussions in Roberts et al. (2003)
It is far more likely to assume mixed results with respect to therapeutic outcomes.

The “go or no-go” decision (Roberts Jr. et al., 2003) will take into account the potential market size and physician adoption with a given efficacy, the pricing and reimbursement options that will exist given that market’s payer situation, and the expected remaining—and refundable—patent term, all weighted by the likelihood of FDA approval given results in the final clinical trials similar to those preceding. Larger markets are therefore defined by attributes of the illness targeted, the safety of the therapeutic, and the efficacy. Larger markets are known to exert significant pull on R&D (Acemoglu & Linn, 2004). It is entirely possible that a therapeutic which has measurable clinical benefit will not see enough of a market to justify further development. Colloquially termed the “Valley of Death” (A. K. Rai, Reichman, Uhlin, & Crossman, 2008), the financial pressures force many projects to be “shelved” at this point.

Phase III trials are a mammoth undertaking requiring 1,000-6,000 volunteers to enter into the program (BIO, 2006). The aim of Phase III, and the driver for enrolling so many subjects participants, is to prove the safety and clinical efficacy of the product over long-term use, incorporating the dosage and safety profile observed in the smaller patient populations in a representative sample of the predicted market. Often, two Phase III trials are needed to sufficiently test the drug. The process lasts, on average, three to four years for classic SMD development. (Y. Friedman, 2008).

Phase III ends with the New Drug Application (NDA). The actual application, prior to electronic filing, required literal truckloads of files to be delivered to the FDA. Again, the goal of the files is to support the marketing claims on the drug’s labeling, as it is these claims which the FDA is ultimately approving or rejecting. The NDA submission also marks the end of clinical trials period as measured from the day of IND submission. The drug is now in its review period.

---

20Phase III trials are the category of trials usually implied with the vernacular use of the term “clinical trials.” Phase I and Phase II can often be populated with known patient populations at health centers. Phase III generally requires larger recruitment processes and hence the common usage.
The review period includes an advisory board of experts which convenes to discuss the merits of the application. They give an up- or down-recommendation. Regardless of their recommendation, the FDA proper has the final say on approval. The focus remains on safety, despite the statutory requirement to prove efficacy. The FDA will not disapprove because more effective options exist but the new drug’s claims must be effective in treating the claimed indications. Under a natural preference for caution, non-treatment might outweigh treatment with a drug of limited efficacy. That is, the final review can be seen as a cost/benefit analysis.

There is some weight implicitly left on the economic incentives preceding the NDA. A safe but minimally-erefficacious product may not create the necessary market depth to justify R&D investment (Dudzinski, 2005). All in all, there remains some regulatory uncertainty at the time of the NDA submission. It is not an all-or-nothing decision, however. The firm may respond to aspects of the FDA advisory board’s recommendation, updating analysis and using their data to address the emergent concerns.

This overview highlights why the New Drug Application process is the longest and most expensive route to approval. In fact, even after approval, there are more costs. There are "Phase IV" clinical trials, which are post-marketing quality control studies that monitoring the now even larger patient pool. These are not merely additional costs to the firm, however, as, in addition to safety checks, Phase IV follow up studies looks towards new uses for the drug (Hope, 2008, p. 57). While an approved drug can be used for any purpose a doctor deems necessary—limited by the laws governing the practice and malpractice of medicine rather than any FDA regulation—the innovator cannot promote a drug for an indication which emerges after initial approval which was not discussed and supported in the NDA. The "off-label" use of a drug is often lucrative for the firm, but dangerous to exploit as promotion or advertising in non-approved directions can result in fines.22

21 For criticisms of potential misuses of the Phase IV trials, namely using the data to convince doctors to use the drug for off-label uses without going through the cost of full clinical trials for that use, see (Angell, 2005, p. 29-31).
22 It is illegal for the pharmaceutical to promote the drug for off-label uses, regardless of how well established
Naturally the Phase I-III segmentations are arbitrary ways to structure the NDA pathway. They developed based on the available science and understanding of therapeutic pathways. Preclinical and Phase I studies slowly and cautiously move an interesting chemical off the lab bench and into animals and then humans, watching for adverse and unanticipated affects in small populations. With changes in science come better predictive models though. Computer modeling of biologic pathways, sometimes referred to in silico testing, have the potential to augment or even replace in vitro and in vivo testing. Some progress here hangs on the willingness of the FDA to accept theoretical- and modeling-supported proofs rather than traditional animal- and human-studies. Whatever changes emerge, the patent narrative is impacted through the changes in cost of the trials. Changes in science therefore affect the parameters under consideration in optimal patent policy (Eisenberg, 2002).

As of yet, the biotechnology paradigm has not significantly altered the average clinical trials approach in terms of length or cost. The most recent biologics legislation however grants biosimilars a shortened pathway similar to that allowed for small-molecule generics. The new pathway is ad hoc, which implies first and foremost a new requirement on the FDA to deal with new information and new techniques on an individual basis. Currently, the expectation is that the need to deal with idiosyncratic applications will open the review landscape to alternative modalities of scientific proof. The rapidly evolving genetic sciences should allow for more experimentation, modification, and human-centered fine-tuning than the organic chemistry of SMD development (Vernon et al., 2006).
All of this becomes relevant to the inquiry here when brought back under the patent narrative's umbrella. There, it was the threat of identical-copies, entering the market by free-riding on the innovator's discovery at new low costs to entry and production, which eroded the ex ante inventive incentives. This drove the SMD regulatory framework and eventually led to the HWA. Although the costs to copy an innovator were presumed low, and the marginal costs of production were equally small, generic copies should have entered the market after patent expiration. Prior to 1984, however, innovators did not see much generic competition due to the regulatory entry hurdle remaining as high for generic copies as for innovators (H. Grabowski & Vernon, 1996, 1986). The HWA therefore lowered the regulatory barriers for generics, provided they were bioequivalent to the innovator, in order to adjust the post-expiration pharmaceutical markets towards the competitive conditions assumed in the patent narrative.

We discuss this simplified entry—the Abbreviated New Drug Application (ANDA)—next. The discussion highlights both the existing incentives and the emerging concern that biologics and biosimilars may respond quite differently to an analogous regulatory situation. Where costs to copy are non-negligible, as they are currently with the characterization and manufacturing difficulties in the biologic-space, a regulatory framework with lowered barriers to follow-on entry may fail to create competitive entry. Here, we have another emerging question too: would-be competitors may prefer to incur extra costs to innovate "above" or "around" the original, securing their own market and competing as substitutes to, rather than copies of, first-generation therapies.\textsuperscript{24} With this in mind we continue the discussion of the regulatory system.

4.1.2 The ANDA and the Hatch-Waxman Act

The second pathway to the market for a pharmaceutical is a shortened version of the NDA, the Abbreviated New Drug Application (ANDA). This path utilizes the safety data

\textsuperscript{24}For an introduction and expansion on the theme, see Grabowski, Cockburn, & Long (2006).
generated through the original NDA process, allowing similar drugs to enter the market on the established safety profile. The ANDA became part of the regulatory landscape in 1984 under the Hatch-Waxman Act. As such, it is often discussed as the Hatch-Waxman pathway. Since the enactment of the HWA, the non-NDA/ANDA pathways have seen more limited use. As the BPCIA follows the NDA/ANDA framework, a focus on these two pathways is generally warranted for economic models.

The HWA aimed to address two problems unique to pharmaceutical innovation relative to other R&D intensive industries. First, it looked at the substantial loss of patent life caused by the competitive need to patent early, before the drug’s market approval. In contrast to other sectors who enjoy a period of market power more closely aligned with the statutory patent life (SPL), which is currently 20 years from the date of patent filing, pharmaceuticals see a shortened effective patent life (EPL). For a patent term starting from the firm’s IND filing, the EPL is the statutory patent term minus the clinical trials period. Calls for redress came as empirical studies found the EPL plummeting from an average 13.6 years to 9.5 years, of the then-SPL of 17 years, between 1966 to 1979 (Grubb, 2005, p. 157-8).

The HWA solution was patent term restoration (PTR). PTR restored a portion of the term eroded by clinical trials and is discussed in depth in Chapter 5. Knowing the patent narrative’s focus on balancing private inducements with social welfare, the HWA also addressed problems for generic entrants post patent expiration. The HWA lowered barriers for generic firms by allowing them to piggy-back their approval on the innovator’s clinical studies. This is the ANDA pathway, abbreviated in that the clinical portion of approval is now seen as complete.

Access to this approval path is clearly predicated on the existence of an innovative product which has completed trials and has a successful period in the market behind them.

---

25 Specifically, Title II of the HWA.
26 The legislative history is the subject of Chapter 5.
27 Title I of the HWA.
As the innovator approaches the end of their newly-restored EPL, an ANDA applicant files for their own approval, based on the safety and efficacy of the original NDA, by showing that they are bio-equivalent (Mandel, 2006, p. 9). Bioequivalence does not mean the generic follower is completely identical, however. Differences between generics and their original siblings may exist. The differences, however, must “make no difference.” Bioequivalence establishes there is a high expectation that innovator and the generic versions will have the same therapeutic outcomes. They both address the indications approved by the original marketing application. Bioequivalence studies are submitted alongside administrative protocol checklists and packaging, handling, and labeling certifications (Mandel, 2006, p. 9). Interestingly, the FDA may not ask for any more safety data than the bioequivalence results (Mossinghoff, 1999).

The immediate effect of the lowering of regulatory barriers for generics was the creation of a generic industry where previously there was very little follow-on entry (H. Grabowski & Vernon, 1996). By using the innovator’s data, the ANDA reduces the NDA-period by half. This is not negligible time, as it still requires the ANDA applicant to manage two to five years of bioequivalence studies and administrative protocols covering manufacturing and labeling (Y. Friedman, 2008, p. 160).

Arguably the most important aspect of the entry smoothing is that starts the ANDA approval process before the innovator’s drug comes off patent (Mandel, 2006, p. 9). This allows generic entry to dovetail with patent expiration, even though the ANDA process

---

28The regulatory pathways utilize the term “same” over “identical.” See FDCA §505(j)(2)(C)(i), §505(j)(3)(D)(ii), cf. (Dudzinski, 2005, p. 195, n.385) See also previous discussion of nomenclature supra at p. 84


31Also see Reiffen (p. 38, 2005). As with the patent system, we abstract from administrative efficiency questions here, but note that administrative benefits to the ANDA include a reduced regulatory load on the FDA. The ANDA review period cuts the estimated time to approval to approximately 19 months (Morton, 1999). There is however significant year-to-year variation in the time to approval and this number stems from the period directly after the enactment of the IWA, 1984-1994, and therefore may reflect some initial startup uncertainty in the application of the pathway, as well as the ramp-up in generic production companies.

does consume non-negligible time. Though there is some contention between innovators and generics, the generic firm's work on the ANDA is generally free of concerns of patent infringement.\footnote{Also notable as a side note is that non-drug uses of the chemical are free to enter the market at the end of original patent term, even though patent-extensions may be in force. That is, a company that found a way to use a chemical as a photographic element would be free to do so at the end of the 17 (or 20) year original patent term, while any generic drug maker could not enter the market until the expiration of the extension periods (Flannery & Hutt, 1985, p. 303). The example comes from Kodak lawyers active in the creation of the legislation and are therefore sometimes still referred to as the Kodak amendments (Lourie, 1984). See also discussion in Lourie (p. 554-560).}

Also notably from the perspective of the recent biosimilars regulation, the ANDA pathway can be employed for generic versions "different from a listed drug; the product can be a new dosage form, new strength, or new route, or a change in a combination product of one of the active ingredients to a different one that is listed in the same class (Fordyce & Cahill, 2002, p. 325-26)."\footnote{Though not debated here, Scherer (p. 32 2007b) relates that after the HWA was passed a group of innovator firms led by Hoffmann-LaRoche fired the president of the Pharmaceutical Manufacturers Association, who had helped make the Hatch-Waxman compromise. The action leaves an impression that innovators may not have been optimistic about the new landscape. Scherer goes on to show that industry profits began an upward trend before and through the passage of the HWA, likely owing to rational drug design and the spread of health insurance, both factors external to the change in regulatory character. Claims that the HWA was successful generally rely on other metrics, including a comparison of the size and scope of the generic sector before and after 1984. See, especially, Grabowski & Vernon (1996,1986).} Though not often mentioned and infrequently used, the ability to modify the follower's product while remaining within the abbreviated pathway ties in with the conceptual framework established for biosimilars. The ANDA set a precedent that may find much more use in the biosimilar world and continues a tradition of importing parts of drug regulation without significant thought to changes in the underlying science.

Nevertheless, the general consensus after nearly three decades is that the HWA was successful (Frank, 2007).\footnote{21 CFR §314.93} The supporting figures are rather stark. Before enactment only 35% of drugs saw generic competition; today almost all innovators find competitors waiting at the patent's expiration (Rouhi, 2002). More specifically, 11 of 13 major drugs had generic entry within 2 months after patent expirations occurring between 1990-1993; in contrast, only 2 generic entrants emerged to compete with the top 13 drugs whose patents expired
between 1976-1982. In sum, costs to enter the market dropped to an average $1-$2 million, translating to price reductions of up to 80% over monopoly prices for consumers (Engelberg et al., 2009; Reiffen & Ward, 2005).

One must stop short of calling the HWA an unmitigated success, however. There are several anomalies in the administrative process that lead to questionable incentives in the post-approval world, especially with regard to incentives to litigate. As the focus in this research is on the ex ante incentives achieved by the combination of patents and PTR, the ex post situation has less direct relevance. The amount of literature here however necessitates a more detailed discussion on the operation of the ANDA process than was required for the NDA.

The Orange Book

The HWA added another feature to the ANDA which was intended to structure information in the now-connected spheres of patent terms and regulatory approvals. Under the HWA, the FDA was charged with keeping track of the valid patents on approved medicines. The Orange Book—formally, Approved Drug Products with Therapeutic Equivalence Evaluations—appeared in 1979 as a simple list of approved products and therapeutic equivalents (Gelber, 2008) but expanded to include patent information. The Orange Book is a list of all the patents an innovator sees as relevant over the approved drug. In addition to simple bookkeeping, the list is a warning to competitors as to what the innovator will view as infringement under generic entry approval.

To be clear, the ANDA applicant is responsible for telling the FDA whether or not there is a patent problem with the generic’s (abbreviated) approval. Generic entrants are left to “certify” the status of the innovator’s patents which are listed in the Orange Book by

---

37 For a complete history of the Orange Book, see Engelberg (p. 402, 1999).
38 The Orange Book can be seen as a canned “Freedom to Operate” Analysis (FTO). FTO analysis is often the first step a company makes in their course of innovation into a space with intellectual property. And FTO maps the known intellectual property rights which may impact upcoming work (Hope, 2008, p. 43-5).
choosing one of four statements for each patent listed by the innovator (Mossinghoff, 1999). A *Paragraph I* statement tells the FDA that the innovator’s drug has not been patented and therefore the ANDA approval has no patent issues.\(^{39}\) A *Paragraph II* states that the patent has expired and again has no problems for the approval. *Paragraph III* certifies a forthcoming date on which existing patents will expire. This lets the ANDA-applicant begin approval proceedings while effectively promising (by admitting the validity of existing patents) they will not utilize the approval until that date. Finally and most importantly, *paragraph IV* certification claims that a listed, unexpired patent is not infringed by the generic entry, or that the listed patent is invalid. *Paragraph IV* statements are therefore viewed as a challenge, admitting that the patent exists and is in force but should place no restriction on the approval of the generic.\(^{40}\)

In a perfect patent system, there would be no ambiguity here. Innovators have however an incentive to list in the Orange Book any and all patents touching on the drug’s sphere of intellectual advancement. This includes patents which do not effect the safety or efficacy of their innovation (Rouhi, 2002; Engelberg, 1999). The FDA has no power of oversight, either on the listing itself or on the content of the patent. That remains the purview of the USPTO. The FDA only compiles the submissions into the Orange Book (Rouhi, 2002).\(^{41}\)

In addition to benefits of appearing as though they control more intellectual real estate than they actually might realistically control, there are regulatory benefits to over-listing

\(^{39}\) Or, that all patents which were applied for were not submitted to the FDA as part of the NDA.

\(^{40}\) A Paragraph Certification is issued for each patent listed in the Orange Book, with the exception of a “section eight” exemption. Here, the generic company states they are not seeking approval for a particular use covered by a particular patent in the Orange Book. The historical rationale was for a generic company to disclose their intent in moving to market (Lourie, 1985c). Section Eight can be seen as an attempt to put the most incentivized agent in the position to understand the often blurry patent landscape and to incentivize their challenge to it.

\(^{41}\) Rouhi (2002) also notes that listed patents cannot be unlisted, unless done voluntarily by the submitting party. Also notably, the Orange Book is not a list of current owners of patents and should therefore not be used as a guide for acquiring licensing rights. The original owner or owners need not record transactions in which the ownership of the patent changes hands (35 U.S.C. 261). This is in stark contrast to the record-keeping of real property. Further, a single property right may have been split among several licensees. Again in contrast to real property, when the assignment of the bundle of rights contained in the patent changes hands (37 CFR 3.1) no one need make a public record of the new deed. An assignment may be recorded according to 37 CFR 3.54 but it need not occur to be a valid transfer of property interests.
one's patents. The HWA allows innovators to file a 30-month injunction against a generic entrant alleging a Paragraph IV certification.\textsuperscript{42} Listing more patents leaves that option on the table and likely contributes to the proclivity to expand a patents' claims in the beginning of a drug's life too.\textsuperscript{43}

The 30-month injunction is the topic of much consternation (Derzko, 2005).\textsuperscript{44} The injunction is itself automatic in that filing the complaint triggers the hold. After the 30-month stay goes into effect, only rectification of the problems in Court can allow the ANDA process to move forward. This is a long period of automatic exclusivity added for the innovator, with little chance of the Courts resolving the issue within 30 months. The Paragraph IV protocol was however implemented to empower the generic entrant to challenge weak or over-stated patents and enter the market as early as possible. It was a bounty for followers, keeping them vigilant of innovator overstating.

To that end, there is a 180-day market exclusivity provision awarded to the first generic entrant to challenge a listed patent under Paragraph IV.\textsuperscript{45} Here, the generic has a half-year duopoly situation where they split the market with the innovator's branded product. It has become the period where the generic firm recoups most of their own sunk costs (Liang, 2007, p. 38). In addition to testing weak patent claims, the bounty coaxes generics into more aggressively demonstrating bioequivalence and moving to market. Exclusivity is not seen as a requirement for generic entry; generics often continue to enter after the initial exclusivity period, especially within "blockbuster" markets (Helm, 2007).

Most recently, the automatic injunction has led to the fiction of "branded generics." Although challenged in Court as anti-competitive, the original patent grant has been interpreted to allow for innovators to reach settlements with the potential entrant. The

\textsuperscript{42}Any judgement by the FDA is stayed for 30 months because the generic company filing a Paragraph IV is, in fact of law, infringement. Therefore patent law takes precedence and the grievance becomes an infringement suit (Mossinghoff, 1999). (35 U.S.C. §271(e)(2)).

\textsuperscript{43}Early versions of the legislation aimed to have the injunction secure only 18 months (Mossinghoff, 1999).


\textsuperscript{45}The bounty exists for Paragraph IV certifications on the ANDA only.
agreements ask the generic to delay entry for a direct payment from the innovator or allows them to be a "branded generic," operating on license with the innovator as a duopoly (Reiffen & Ward, 2007).46

Though these incentive problems are rather obvious outcomes of the legislation's wording, we can explain them as compromises forged during an inopportune time in the HWA's debate: the Roche v. Bolar reversal. Before the reversal, a legislative compromise had been worked out between pioneers and generic firms regarding experimental exemption for generics to begin working toward manufacturing and approval (Engelberg, 1999). After the legal reversal, the compromise seemed unwise for pioneers and they changed their stance in the debate. A new compromise had to be created quickly to salvage the broader structure of the HWA's balancing (Engelberg, 1999, p. 402). Anecdotes like this explain many of the HWA's less-efficient outcomes, and a full history of the legislation is taken up by Chapter 5.

Without wading farther into the debate, it is clear that incentives in the ex-post world less aligned than efficiency would dictate. The 30-month injunction has become an expensive event for a generic firm (Liang, 2007, p.39) and has engendered some rather questionable legal arrangements between sides. There have been small attempts to control moral hazard.48 While any empirical study of the issues is beyond the scope of this project, the issues are important to the pressure they have put on any future legislative attempts to adjust the pharmaceutical market.

46 This has also been looked into by the FTC. See Geneva Pharmaceuticals, Inc. & Abbott Laboratories, Analysis to Aid Public Comment, at http://http://www.ftc.gov/os/2000/03/genevaabbottanalysis.htm. [Last visited July 12 2008], and on file with the author. See also discussion in Rai (p. 184, n. 49, 2001).
47 733 F.2d 858, 221 U.S.P.Q. (BNA) 937 (Fed. Cir. 1984)
48 Congress limited innovator firms to a single 30-month stay and removed the ability to stack such claims with the Medicare Modernization Act of 2003. Pub. L. No. 108-173, §11101 (Dec. 8 2003). 117 Stat. 2066 (2003). The FDA also moved to remove ability to stack multiple 30-month stays. 68 Fed. Reg. 36676 (June 18, 2003) (amending 21 C.F.R. sec. 314.95). See discussion in Liang (p.39, 2007) and Epstein & Kuhlik (2004). Although recently prohibited, innovators used to wait to file suit, allowing damages from infringement to accrue before acting (Corbit, 2008). Firms must now also move to sue infringing entrants within 45 days or they lose the ability to file a grievance. Also, the U.S. District Court found, and the Court of Appeals upheld, the ability for the courts to extend or shorten the 30-month stay in response to parties that may be uncooperative in the trial process. See Eli Lilly and Company v. Teva Pharmaceuticals, US Court of Appeals for the Federal Circuit, 2009-1071.
4.1.3 The 505(b)(2) NDA

The third pathway to FDA approval, the §505(b)(2) is available to new SMDs and to biologic applicants.\textsuperscript{49} Dudzinski (p. 198, 2005) summarizes the §505(b)(2) pathway as a hybrid residing between the NDA and the ANDA. It expands on the pre-existing Paper NDA pathway.\textsuperscript{50} The §505(b)(2) is useful for changes in approved drugs where the modification is too broad for the replication-oriented ANDA route but enough data exists to bridge existing safety information to the change.\textsuperscript{51} The path is most notable in that it has received attention in the recent biologics and biosimilars regulation.

Similar to the ANDA route, the 505(b)(2) pathway requires applicants to disclose what patents their application may impact.\textsuperscript{52} Unlike the ANDA, however, there is no requirement for bioequivalence and therein lies its broad applicability, especially to biologics (Dudzinski, 2005, p. 198). In fact, true generics cannot be reviewed under the §505(b)(2). Such use would take its focus away from determining whether the existing safety data supports a modified drug’s approval (Liang, 2007, n.184).\textsuperscript{53} The case law illustrates a willingness by the FDA to extend its own interpretation of what constitutes “sameness” in different scientific contexts here (Dudzinski, 2005, p. 203,4), a willingness to weigh the particular ramifications of a drug now mirrored in the BPCIA.\textsuperscript{54}

A general overview shows that the amount of innovator data available, and the ability for applicants to build bridges to that data, is not unbounded (Dudzinski, 2005).\textsuperscript{55} Dudzinski (p. 210, 2005) summarizes the dynamic between inventors and generic firms within the

\textsuperscript{49}As of 2005, both Follistim® by Organon and GlucaGen® by Novo Nordisk, had successfully used the §505(b)(2) pathway for their biologic drugs (Dudzinski, 2005).

\textsuperscript{50}See for a complete discussion, Fordyce (2002)

\textsuperscript{51}A similar dynamic exists in the pathways for medical device approval.

\textsuperscript{52}FDCA §505(b)(2)(A)-(B) (2000)

\textsuperscript{53}Note that identical copies would be reviewed under §505(j)(2) of the FDCA (Tucker et al., 2008, p. 58)

\textsuperscript{54}Established firms have noted that expanded use of §505(b)(2) could amount to a government taking of confidential information, submitted in open format to support an innovator’s NDA (Dudzinski, 2005, p. 206). Others find that the current usages do not amount to takings (Yoo, 2005). There was an analogous debate at the time of the IWA’s enactment and the contemporaneous Bolar discussions as to whether a Bolar-exemption would be a taking (Engelberg, 1999, p. 404-405).

\textsuperscript{55}See esp. Pfizer Inc. v. Dr. Reddy’s Labs., Ltd., 359 F.3d 1361, 1366 (Fed Cir. 2004)
§505(b)(2) pathway as a “fine line,” where the generic must achieve a “threshold similarity so that the generic can utilize section 505(b)(2) and rely on the pioneer’s studies, but avoiding infringement of the pioneer patent.” While a deeper discussion is not necessary for modeling of ex ante patent incentives, the pathway is notable for its foreshadowing of the biologics regulation (Dudzinski, 2005).

4.1.4 The Paper NDA

The final route to market approval is the so-called Paper NDA. The Paper NDA is similar to the §505(b)(2) path and the two are often seen as “two-general types of section 505(b)(2) applications (Dickinson, 1999, p. 196).” The basic idea here was to allow applicants to cite the published paper of others to establish the safety of their drug via proof existing in the public domain. The bridge here is not to an innovator’s trials data but rather the general but persuasive body of scientific discourse. Supporting information was supposed to come from academic, published sources (Fox & Bennett, 1987, p. 95).

While another lightly-used approval pathway, with most of its use predating the HWA, its existence highlights the limited data sharing which currently accompanies the regulatory system. Data sharing in the ANDA exists only at the FDA approval level. The Paper NDA shows a willingness on the part of the regulator to consider external information as a replacement or supplement for clinical trials while highlighting the lack of data sharing that actually occurs. Using academic publications is quite sensible in theory—especially when considering the amount of research that is funded by public sources—if less utilized today. The availability of public sources inclined SMDs in well-established areas to enter with new formulations, areas and similarly to bring changes to previously approved SMDs to market. These changes generally included modifications of dosage, strength, and routes of administration (Glover, 2007).

There are however practical concerns. Pharmaceutical firms have long espoused conflicts with publishing. Just as with patent disclosure, it is often an advantage to bring in fresh
insight and research experience. The advantage is there until a solution is found, and then the academic inclination to publish finds itself at odds with the competitive advantages of secrecy (Lesch, 2006, p. 54). The FDA, as a federal agency, follows statutes barring federal employees from disclosing trade secrets.\footnote{21 U.S.C. §331(j) and 5 U.S.C. §552(b)(4) and 18 U.S.C. §1905. (Dudzinski, 2005, p. 214, n. 529)} Secrets can be revealed or implied by publications, however, and that leakage is not protected by regulations. The disinclination to publish important information may be the cause of the Paper NDA’s limited utilization. By 1984 the FDA estimated that sufficient public sources were not available for 85% of post 1962-drugs which could have benefited from the pathway (Fox & Bennett, 1987, p. 97). In the end, the Paper NDA only succeeded in moving a few approvals into the market being functionally superseded by the §505(b)(2) pathway (Dudzinski, 2005, p. 220).

While relevant in theory and carrying interesting potential into the information-rich biotechnology age, the Paper NDA route needs modern clarification. The §505(b)(2) clarified some issues for the SMD market and thereby eclipsed the academic route. Nevertheless, it again shows an indication of Congress’ desire to allow abbreviated entry wherever possible. A modern revision to the paper-routes, employing science and established resources rather than costly trials, could have significant advantages for society.\footnote{For a more complete discussion including legislative history, see (Dudzinski, 2005, p. 214-16)} The existence of these secondary approval routes should therefore inform the reader of the regulatory landscape’s proclivity toward allowing entry where proof exists, even when such proof takes forms varying from the clinical trials form.

\section{4.2 Biologics Regulation: The Emerging Landscape}

The regulatory landscape, dominated for the past three decades by the NDA and the ANDA routes, is changing. Thankfully for research in innovation policy, we are changing in the directions in which previous and overlapping regulation already inclines the industry. This lends some inertia, in addition to past precedents, to our predictions.
Most notable for the research at hand, the landmark health-care legislation signed into law in March of 2010\textsuperscript{58} contains a section codifying the long-discussed abbreviated approval process for biosimilars.\textsuperscript{59} Title VII, Subtitle A, of the Act—\textit{Improving Access to Innovative Medical Therapies: Biologics Price Competition and Innovation} (BPCIA)—sets out for the first time the process by which producers of biosimilars can capitalize on the data and clinical trials conducted by pioneers. The section is the biologics' answer to Title I of the Hatch-Waxman Act.

Biologics needed a personalized answer because they have resided under a different approval pathway than SMDs. Biologics are regulated under the Public Health Services Act (PHSA), which predates most of the FDA's chemical regulation.\textsuperscript{60} This existed because of the early used of serum therapy and the harvesting of insulin and vaccines from biologic—and often non-human—sources. Such development took a backseat to chemical therapies for the majority of the modern pharmaceutical industry's development. And because the biologic approval path was not as well-worn and litigation-tested as the SMD routes, the debate leading up to the BCPIA's enactment rekindled a discussion on optimal innovation policy within drug development not seen since the HWA.

The FDA defines a biological product as a "virus, therapeutic serum, toxin, antitoxin, vaccine, blood, blood component or derivative, allergenic product, or analogous product or arsphenamine or derivative of arsphenamine (or any other trivalent) organic arsenic compound, applicable to the prevention, treatment, or cure of a disease or condition of human beings."\textsuperscript{61} The definition lacks an auxiliary definition for generic biologics but it is clearly separate from the machinery for chemical drug regulation. SMDs fall under the regulations provided in the Federal Food, Drug, and Cosmetic Act (FDCA) (Karst, 2004).  

\textsuperscript{58} \textit{Patient Protection and Affordable Care Act} (PPACA), Pub.L. 111-148, 124 Stat. 119.


\textsuperscript{60} 42 USC Chapter 6A.

\textsuperscript{61} PHSA §351(i).
Although regulated differently, biologics remain under the FDA’s oversight for market entry. Approving a biologics requires a biologics license agreement (BLA), which is the analogue to the NDA (Karst, 2004). As a “Biologic-NDA”, the BLA is a complete approval pathway which includes clinical trials.\textsuperscript{62} And despite the separation, the process of approval for each is nearly identical (Mandel, 2006, p. 12).\textsuperscript{63} If a BLA is issued for a biologic, no NDA filing is necessary (Dudzinski, 2005, p. 184).

As noted above, many ideas from the regulation and approval of pharmaceuticals were imported in biologics regulation, both proposed and enacted. Also, as noted in Chapter 3 there is much overlap in biotechnology and small molecule development and bright-line boundary dividing large- and small-molecule therapies. Certain biologics whose pharmacological properties are enough to their chemical cousins have entered the market under FDCA regulation (Liang, 2007, p. 39-42). While ostensibly predicated on the relative simplicity of the biologic itself, the fact led to much debate by lending biosimilar manufacturers a passable resemblance to SMD generics. Interested parties noted these examples as an indication that biosimilar entry is safe; characterization and manufacturing problems were obviously surmountable.

Entering US law inside such a substantial rewriting of the US health care system, and notably one which drastically impacts government payer programs like Medicaid and Medicare, cost savings from generics of all types was viewed favorably by the PPACA’s proponents. The question of how to bring biosimilars in line with generic drugs created a long debate leading up to the inclusion of the BCPIA in that legislation. Though it is too lengthy to be covered in depth,\textsuperscript{64} the general outcome is a framing, but vague, structure. It allows the FDA much room to define the entry hurdles as technology warrants. The major themes in the regulatory framing have all been touched on here.

\textsuperscript{62}See Carver et al. (2010) for a discussion of the administrative duties in the FDA with regard to biologics.
\textsuperscript{63}Mandel (2006) implies that it is almost as if the FDA had worked consciously to make them as similar as possible. (61 Fed. Reg. 2733, 34-36 (1996)).
\textsuperscript{64}See Carver et al. (2010) and Morgan (2010).
Entrusting the FDA to regulate biosimilar entry in an ad hoc manner, framed within the broader safety and efficacy mandate, may be a significantly less impactful choice than the analogous regulatory changes which accompanied the HWA’s generic entry provisions. From the perspective built from the preceding chapters, layering patents, industrial organization, and the regulatory hurdles of entry, the BPCIA’s new regulatory exclusivity provisions are potentially monumental changes.

Exclusivity is not an entirely new addition to the pharmaceutical sector. The magnitude of exclusivity granted under the new regulation is however a marked departure. The HWA grants 5-years of FDA data exclusivity to an approval, as long as they contain chemicals that have never before been approved (Y. Friedman, 2008, p. 153). This means that irrespective of the patent profile, the FDA will not approve an ANDA based on the innovator’s data within 5 years from the innovator’s approval. Though the patent may expire in that period, a generic copy would have to supply their own data should they wish to enter the market within the 5 year span.

This existing exclusivity places an effective floor for the expected market power granted to new chemical entities (NCEs). In aggregate, this turns out to be 5 – 7.5 years of regulatory exclusivity, where 5 years come from this statutory data exclusivity provision and the extra 2.5 stem from the 30-month “challenge stay” discussed in Section 4.1.2. Even without an effective challenge, it was believed at the time of the HWA’s enactment to take 2 years to approve the new generic drug, thus giving even a patent-less chemical on the day of their approval 7 years of effective non-patent exclusivity (Engelberg, 1999). It has only been the general case that patent terms, with restoration, extend past exclusivity and so they have born the brunt of the empirical analysis. The underlying exclusivity provisions however provide a safety net or lower bound to the term of market power granted to innovators.

The biologics situation did bring new fodder to the debate. Of note is the distinction

---

65 The Drug Regulation Reform Act of 1979 included non-patent exclusivity similar to those eventually placed in the HWA, and 1982 had seen the Orphan Drug Act, which grants exclusivity based on indicated disease (Fox & Bennett, 1987).
between data exclusivity and market exclusivity. The former, active in the HWA's 5-year regulatory exclusion of ANDAs, protects the use of the innovator's original data. The later is more powerful and grants exclusivity over a therapeutic's market area. Market exclusivity's power comes through the gatekeeper's ad hoc ability to assess an impending competitive situation within a therapeutic market. Considering only the patent protection profile, a non-infringing therapeutic with its own clinical data may enter the market immediately as the FDA need not consider the patent situation outside of the NDA/ANDA arena. This is an illustration of the common semantic retort a patent is not a monopoly grant. A patent certainly creates the potential for a monopoly but the ability for competitors to invent around any original intellectual content such that their invention plays in the same market means non-infringing competition does occur within patent-structured product markets.

For pharmaceuticals, a non-infringing therapeutic would have to generate their own clinical data to enter around the FDA-granted data exclusivity, but could arguably still enter. Regulatory exclusivity, on the other hand, ostensibly bars the FDA from approving a non-infringing therapeutic with proprietary clinical data should that data support an application for treating the same indications as the original. Exclusivity of both kinds relies on the gatekeeper to disallow the competitor into the market, with data exclusivity barring a competitor from referencing a pioneer's safety data and the market exclusivity preventing overlap in treatment markets. The latter claim is largely untested, though the exclusivity provided under the Orphan Drug Act does give the FDA some experience here.

The debate preceding BPCIA focused on providing exclusivity provision rather than changing or strengthening the PTR provisions. The growing price tag of developing biologic medicines, combined with the uncertainty as to how biosimilars may proceed to the market, was seen as placing large requirements on the patent system to guarantee profits. The requirement for patents to secure profits is increasingly uncertain, where the science of

---

66 The Orphan Drug Act is a current example of market exclusivity.
biotechnology is seen as inherently better suited to inventing-around intellectual content. Further, there is a trend and perception of biotechnology patents being interpreted narrowly in court (Rimmer, 2008), enabling followers to invent-around at a closer radius to the original, likely becoming a closer substitute in the therapeutic market.

During the debate, Grabowski et al. (2008a) found that biologics need to absorb profits for 12 to 16 years to justify the cycle of investment which funded the existing biologic drugs. 12 years of exclusivity is indeed what Congress chose to implement. Brill (2008) however used a more cautious set of parameters and argued for a significantly shorter break even times.\(^{67}\) Henry Waxman’s own bill\(^{68}\) asked for only 5.5 years of exclusivity. His HWA compatriot Sen. Hatch however co-sponsored the amendments for 12 years of exclusivity.\(^{69}\) While all sides issued rhetorical arguments on the pros and cons of the proposals with regard to social welfare and growth, the legislative focus was clearly directed on adequate remuneration.\(^{70}\)

At this point, the conversation should appear as just another incarnation of the social balance debate whose roots extend back to the historical derivation of the patent narrative in Chapter 1 and the original patent length optimizations of Chapter 2. The intervening chapters exploring the Pharmaceutical industry’s structure and its own foundations on their underlying scientific understanding (Chapter 3) and this regulatory introduction illustrate a potential for a tangled situation; one cannot not expect a change in a single innovation policy parameter within the patent/science/regulatory triad to have clear cut effects on the outcome. At the very least, incentives added for the benefit of industry will incur a more complicated calculus in the social sphere, where additional innovation bring entirely new markets and their associated welfare but extend, or create new, detriment from market

\(^{67}\)The Obama administration also sided with shorter times. see http://www.reuters.com/article/healthNews/idUSTRE5506ZZ20090625 [Last Accessed August 2009]; see also FTC, Emerging Health Care Issues: Follow-on Biologic Drug Competition (2009).

\(^{68}\)H.R. 1427

\(^{69}\)in S. 726.

\(^{70}\)A reflection of the HWA’s own history. See Chapter 5.
power structures.

This situation is seen most concretely in the new exclusivity provision overlap on the patent protection afforded to biologics. Although some regulatory and patent overlap has always existed, the new magnitudes shift the balance of protective power to the exclusivity protection. Kotlikoff (2009, 2008) echoed earlier points by Eisenberg (2004) in questioning the role of exclusivity on top of the existing patent incentives. Do the dual systems “over-power” the intellectual property of drug companies at the expense of society? Or is the overlap simply another extension of the patent narrative, where complications created by uncertainty in the drug development process now require more active involvement to assure the same level of R&D incentives enjoyed by non-pharmaceutical projects? The debate is just beginning.\(^{71}\)

The simple continuity of the debate does drive home the fact that optimal innovation policy research has not reached a consensus. The back-and-forth within the biologics and biosimilars discussion further emphasizes the HWA’s implied certification that industries do receive patent incentives in the same way. This all underlines the only minor consensus within research on innovation policy: a “one-size-fits-all” incentive system is unlikely to be efficient. In the modern age, any invention, regardless of its substance, level of intricacy, benefit to society, or cost to develop, remains eligible for the same 20-year protection of their intellectual contribution. In the eyes of the USPTO and global patent law, and stated through the admittedly hyperbolic rhetoric of the system’s harshest critics, building a better mousetrap remains as worthy of 20 years of protection as uncovering a cure for cancer.\(^{72}\)

On the other hand, the one-size-fits-all system has proved remarkably adaptable, developing in the legal sphere the ability to encompass software and genetic patents the way it originally protected mechanical innovations. In this vein, PTR remains entirely unique

\(^{71}\)In addition to questions of overlap, administrative efficiency again enters the discussion, with proponents of exclusivity finding it up to the task of inducing R&D with less waste (Morgan, 2010) or with less uncertainty (H. Grabowski, Long, & Mortimer, 2008).

\(^{72}\)Notably, Posner (2012) and Bessen & Meurer (2008) provide structured introduction to the critical literature.
in its ability to adjust the patent term. Although the particular framework of the HWA restraints PTR to restoration, strictly held to be less than any the time lost to patent term erosion, the general functionality is academically interesting. Furthermore, the IIWA will continue operating for biologic approvals as it did for SMDs. Though the power of exclusivity incentives may eclipse the patent incentives, these earlier policies will provide continuity between the regulatory environments.

The next chapter therefore steps back to (re)introduce the idea of patent restoration. We accomplish this by describing the legislative history of the IIWA. This provides a direct connection with—and review of—the literature on patents in Chapter 2. The legislative debate adds an appreciation of the changing markets and the competitive dynamics illustrated in Chapter 3. That blending led to the two sides of the IIWA: patent term restoration for innovators and lowered barriers for entry for generics. Our understanding of the complications introduced by biotechnology and the recent regulatory updates provide a backdrop to the history, and the combination then completes the narrative on which the modeling of firm incentives and welfare maximization take place in Chapter 6.
Chapter 5

History and Complications of Patent Restoration

Politics is the shadow cast on society by big business.

John Dewey

The Hatch-Waxman Act (HWA) of 1984 was a substantial change to both the intellectual property and regulatory landscape of the pharmaceutical sector. The HWA is first a unique exception to the one-size-fits-all patent system, with application solely to pharmaceutical products which successfully complete regulatory review. Second, it is a major restructuring of the post-expiration competitive landscape, effectively dropping the regulatory burden for generic entrants. Although generally acknowledged as a successful policy (Frank, 2007), the established empirical and legal literature focuses primarily on its ex post effects of fostering price competition.\footnote{Inter alia, Grabowski (1986); Special Issue No. 54, Food & Drug Law Journal (1999); Also Lourie (1984,1985,1985a,1985b,1986); Flannery and Hutt (1985); Mossinghoff (1999); Engelberg (1999); Fox & Bennett (1987); Haddad (1978); American Enterprise Institute (1981).}, Analysis of the restoration framework itself as an ex ante incentive to invest in R&D commensurate with its title of patent term restoration, is nearly absent.

Particularly absent from either conversation is a discussion as to the restoration rate
chosen by the HWA. The PTR rate is the level at which regulation credits patent time lost during clinical trials back to innovators after approval. That rate should be a significant decision when one takes into consideration the role the statutory patent life (SPL) is presumed to play in the patent narrative. Seen from the economics literature on optimal patent incentives, the effective patent life (EPL) determines what R&D projects can justify their investment and which cannot. If the SPL is a global norm and some sector-specific phenomenon causes \( EPL < SPL \), the level of innovation for projects in that sector will be lower than in others. The potential for social welfare deviations between sectors is then apparent. Despite the potential for both enhancing patent incentives and adjusting maximizing social welfare, and with very little recorded discussion, the HWA chose a one-to-two rate, granting innovators half of their clinical trials time back after approval.

Fox and Bennett (1987) note that dearth of published legislative history makes starting an investigation difficult. To be clear, there was no lack of debate but the majority of negotiations were handled off the floor. Though not uncommon, the lack of written documentation that came to the floor effectively kept all conversation out of the public congressional record. This presents investigators with the problem that “[t]o really understand the intent of this legislation, [they] almost had to be there as it was negotiated (Fox & Bennett, 1987, p. v).”

This chapter takes up the task of parsing together a coherent conversation from the various sources. The goal, following previous chapters, is to understand the breadth of issues as perceived by the actors involved. The chapter therefore presents the major points and incentives which Congress, the innovative drug industry, the generic industry, and various interest groups, perceived as crucial. Some points have obvious connections with the prevailing science and industrial structure of the time and find support in the discussions of Chapter 3. Other points, notably including the Act’s due diligence provisions, reveal a

\(^2\text{Key references from those who were there at the time but which were published outside of Congressional records are: Alan D. Lourie (1986,1989) and Flannery & Hutt (1985); as well as William F. Haddad. Finally, note discussion on early bills presented by the American Enterprise Institute (1981).}\)
thread of distrust in the debate which, while plausible, finds little support in our introduc-
tions. The formal model of Chapter 6 can then take all of these topics into account for a
final critique on the IIWA's abilities and inabilities.

5.1 Historical Precedent & Future Pessimism

The concept of restoration of patent terms was not a novel idea at the time of the HWA's
enactment. Though not formal restorations, term adjustments have long been part of the
IP landscape. Their inclusion is a notable reminder of the patent narrative's focus on
rewarding invention, discussed in Chapter 1, especially viewing patents as ex ante providers
of incentive for R&D. When we put weight on the imperative to remunerate inventors for
the social welfare they create, and have already chosen as a society to do so via patenting,
an unnaturally truncated patent term is clear problem.

Historically, Great Britain and the U.S. made provisions for patent extensions for in-
vengers who lost the ability to market their product due to war-time conditions (Desrosiers,
1990). Wars cause both backlogs in patent approvals and change the consumer market for
inventions, with clear impact on the ability to profit (Ochoa, 2002, p. 54). Naturally, the
ability of inventors-turned-soldiers to work their patent is also impacted. The historical
realities have developed into a precedent favoring redress when the patentee is harmed by
government action (Cooper, 1993, p. 72).³ Above legal nuances, the patent narrative is
quite applicable here.

British patent policy also had very strong linkages to the remunerative patent story. In
addition to their war time provisions, they used to allow patent extensions under a showing
that even peacetime circumstances had caused a failure to procure sufficient profits (Grubb,
2005, p. 155); this is no longer in effect, however. In the U.S., Congress retains the power

³see Patent Extension Hearing, Hearings on S. 526, S. 1165, and S. 1506 Before the Sub-committee on
Patents, Copyrights & Trademarks of the Senate Comm. on the Judiciary, 102d Congress, 1st. Sess. 2 (Aug
1, 1991).
to extend patents by formal declaration outside of wartime. This has only been used occasionally (Walterscheid, 2004, p. 601, n. 15). When used, it has always been justified on the public’s benefit. For example, consider the the 19th century’s 7-year extension to the then 14-year patent. The wording of that act specifically justified the extension on reward and remuneration grounds (Mossoff, 2007, p. 1006). The remuneration philosophy is even apparent in the repeal of the extension, which happened when the patent term was adjusted from 14 to 17 years in 1861 (Mossoff, 2007, p. 1004).

Although wartime economic restrictions are a rather stark circumstance on which an inventor may suffer harm, it is not a large legal jump to viewing regulatory burdens in the same light. Today, there remains only the arguably remote possibility of private legislative extension (Cooper, 1993), and the recent (1995) increase of the statutory patent term to 20 years starts the patent from the date of filing, rather than the 17-year grant which started at the date of issue. Thus, an adjustment for a regulatory burden which impacts the inventor’s ability to profit, and thereby would diminish the incentive effects of the statutory patent grant already available to others in unburdened situations, has strong historical ties.

---


6Mossoff (p. 1006, 2007) explains quoting from the 1836 Act: “...if the patentee ‘failed to obtain from the use and sale of his invention [during the fourteen-year patent term], a reasonable remuneration for the time, ingenuity, and expense’ in creating the invention, he could obtain a seven-year term extension.” Patent Act of 1836, at 125. cf. Ochoa (2002).

7Eldred v. Ashcroft 537 U.S. 186 (2003), stating that the term “limited in the Constitution authorizing the granting of patents did not mean “fixed” or “unalterable,” and that the call to promote progress of science was just a preamble and had no substantive voice (Walterscheid, 2004, p. 605). The Supreme Court has not directly addressed the issue of the constitutionality of patent term extensions, but has “decided cases in reliance on such revivals and extensions (Cooper, 1993, p. 61).” For other constitutional issues, including overlaps between the HWA and GATT, see Marks (p. 466, 1996).

8The 1861 change to a 17 year term, measured from the time of issue, was the only major patent term extension until 1995. The 1995 adjustment came under the U.S.’s treaty obligations to the WTO’s GATT treaty. Signed into US law as The 1994 Uruguay Round Agreements Act (URAA), Pub. L. No. 103-465, 108 Stat 4809 et seq. (1994), codified as amended at 35 U.S.C. §154 (Supp. 1995), (effective June 8, 1995). For a complete discussion, see Marks (1996). This increase was simply to bring the U.S. in line with their trading partners and was debated as an adjustment of U.S. innovation policy (Walterscheid, 2004, p. 602).
and a showing of need.\(^9\)

The uniformity imposed by the GATT makes it unlikely to use Congressional exceptions for anything approaching targeted adjustments to statutory patent life. In fact, the GATT limits patent extensions to restorative situations, proscribing almost all industries besides pharmaceuticals. Even here, there are concerns that even HWA-style PTR could violate GATT provisions (Eisenberg, 2003).\(^{10}\) Therefore, and despite both historical precedent and a large economic literature espousing benefits for targeted innovation incentives, the push for harmonized global trading conditions means the HWA is likely to remain the unique exemption to the one-size-fits-all patent system.

5.2 The Legislative History of the HWA

The road to creating a unique patent term provision for pharmaceuticals began in 1978, when a policy review board under the Carter Administration inquired into domestic innovation rates. The goal was to identify areas in which innovations might drive economic improvement and job creation. The Committee found a growing concern in the pharmaceutical industry that patent term erosion was reducing incentives to invest in R&D projects. Erosion occurs because pharmaceutical firms most often patent their findings early in the basic research phase of R&D. Substantial time then elapses long before they approach the FDA for marketing approval. Many of the patent’s effective years are therefore lost during the R&D process rather than providing remuneration for that R&D expenditure.

The erosion problem might have been avoided when the US Patent Office, following 1962’s Kefauver-Harris Amendments, stated that they would not grant a patent on a therapeutic drug until the compound was found safe and effective (Engelberg, 1999). The Court

\(^{9}\)The USPTO did estimate that ninety-three patents which had HWA extensions did receive an “extra” increase with the transition to the twenty year patent (Marks, 1996, p. 461,n. 155).

\(^{10}\)This was seen in a dispute over national (Canadian) provisions which allowed for differential treatment of pharmaceuticals, where a trading member noted that the GATT mandates equal intellectual property protection regardless of industry.
of Customs and Patent Appeals however reversed the proclamation on the grounds that the USPTO’s fundamental goal in issuing patents was to incentivize capital investment necessary for development. The ruling made it legally possible to obtain a patent on drug research; the conditions of competitive industrial R&D made it strategically viable to get those patents early (Engelberg, 1999, p. 394). Patent term erosion was then part of the financial reality of modern pharmaceutical development, and the average drug which had enjoyed 13.6 years of EPL in 1966 enjoyed on average only 9.5 years by 1979 (Lourie, 1984). Erosion combined with other subcommittee findings on the growing complexity of clinical trials, culminating in a recommendation for Congress to create a patent term restoration policy for pharmaceuticals (Mossinghoff, 1999; Lourie, 1984).

It is notable to consider that despite the decline in effective patent life, there was no commensurate decline in R&D (Lourie, 1985c). With wonderful foreshadowing on the biologics debate in Chapter 4, stable investment was credited to the de facto non-patent exclusivity created by the FDA’s requirement that generics repeat clinical trials. Despite acknowledging the role of those barriers, the first bills introducing PTR did have any components for smoothing generic entry, a facet which would eventually define the Hatch-Waxman compromise between innovators and generics.

Early attempts to restore patent life for pharmaceuticals follow immediately after the Carter Administration’s studies but the HWA would not emerge for several years. Compared with the encompassing restructuring eventually created, the early attempts were quite

---

11 As a historical aside, Senator Kefauver had proposed reducing the patent term of drugs due to what he found as exorbitant profits and markups (Desrosiers, 1990). Those provisions were eventually defeated with strong opposition but Kefauver’s name still appeared on the Kefauver-Harris amendments to the FDCA in 1962. The new mandate to test efficacy in addition to safety of new drugs eventually led to the shorter effective patent terms.

12 See also Eisman and Wardell (1981). Hansen (1979) had found a similar decline from 16.5 years in 1960 to 9.7 years in 1978. The University of Rochester found a decline of 13.8 years in 1966 to 8.9 years in 1977 (Desrosiers, 1990, p. 121,n. 59); (American Enterprise Institute, 1981)

13 See Remarks of Rep. Kastenmeier, House Floor Debate, Cong. Rec. of Aug. 8, 1984 at H8708, discussing the parts of the OTA study not in support of PTR.

14 H.R. 3589 (1979) by Congressman Steven D. Symms of Idaho; A Bill to amend title 35, United States Code, to extend the patent term for new drugs and new animal drugs. The Representative’s name is apparently misspelled in Lourie’s (1984,1985) history of the DPR-PTC legislation as “Symes,” but is correct in Lourie (1985b).
targeted. The first bill simply attempted to move the patent grant on a human or animal drug to run 17 years from date of market approval, or 27 years from patent grant, whichever was earlier (Lourie, 1984). Lourie (1984) noted that “[w]hile its goal was laudable, patent lawyers knew that it attempted to deal with a complex problem too simply.”\textsuperscript{15} The initial attempt is most important in setting a limit to the subsequent debate. The focus was to remain on replicating the patent life of other (non-drug) inventions within the pharmaceutical context. That is, this was an effort for restoration and not for extension, optimization, or to support an activity with particularly important social benefits.\textsuperscript{16}

The incentivizing effects of the (then) 17-year SPL were taken as given the task of legislators was to be restricted to restoring that incentive. Remarks made by the Commissioner of Patents and Trademarks illustrate an unflinching attachment to that existing term, committing to supporting legislation that would “provide an equal term, not less than seventeen years of effective patent protection (Marks, 1996, p. 448).” The sticking points to come were therefore not to be drawn from the larger patent optimization literature of Chapter 2. The emerging concern was rather with how pharmaceutical firms would react to the structure of the restoration.

The cause of patent term erosion was clearly the regulatory review period. In addition to bounding the restoration discussions to pharmaceuticals only,\textsuperscript{17} the restorative-focus had

\textsuperscript{15}The entire legislative history is recounted by Fox and Bennett (1987), and Lourie (1985, 1985a, 1985b). Alan D. Lourie served as a member of the Pharmaceutical Manufacturers Association and later as chairman of that committee (Lourie, 1984, p. 527).

\textsuperscript{16}Traces of a desire to optimize patent policy are detectable throughout the legislative debate, though they are somewhat confused. See, House Report Part I, at p. 40-41, included in Fox & Bennett (p. 153, 1987): “The Committee established different maximum periods of extension to provide greater incentive for future innovations. By extending patents […] the Committee expects that research intensive companies will have the necessary incentive to increase their research and development activities.” Further, the stated intent of Title II of the IIWA became to create “new incentives for increased expenditures in research and development . . . ” (Fox & Bennett, 1987, p. 177). Note the variable use of the adjectives “new”, “greater”, and “necessary”, modifying the implied target and magnitude of PTR’s effect.

\textsuperscript{17}The bills decidedly did not extend to “all compounds of the involved patent or claim, and [are] not to include totally distinct types of uses for a claimed product (Lourie, 1984, p. 528).” Of particular note, the petroleum industry put forward in 1984 that environmental regulations were similar to the hurdles placed on pharmaceuticals by FDA regulations. They offered a concrete example of eroded patent terms for new catalysts which were to be used in new refineries. As new refineries had to wait for construction to being while environmental approvals went through, a new catalyst’s effective patent term would be affected similarly
a clear enemy. The regulatory review became the only relevant evidence that a patent's incentives had been sufficiently impacted external factors (Fox & Bennett, 1987, p. 109).

The initial attempt at PTR\textsuperscript{18} was also notable for its effect in creating what was to become the opposition party. The generic industry, as well as interest groups representing retired persons, entered into the discussion. Their first impact to generate more studies, now including the social impacts of patents on pharmaceuticals (Lourie, 1984). The Congressional Office of Technology Assessment (OTA) confirmed in 1982 the findings of the Carter Administration that patent terms had been eroded. The OTA study suggested that extensions could encourage development while noting that profits in the sector had not declined. The pharmaceutical industry was left to defend their pricing decisions while simultaneously continuing the PTR campaign for restored protection.

They did so with appeals to adequate remuneration. They cited the industry's flat sales and rising R&D expenditures (Desrosiers, 1990, p. 128). Unfortunately, the opposing statistics were quite stark, including a 320% increase in average prescription prices between 1974 to 1987 (Desrosiers, 1990). Although the claim for parity with other R&D industries fit with the one-size-fits-all patent system, the social dimensions quickly cemented the opposition's concerns into the same story, forcefully expanding the debate to include access to drugs, pricing decisions, and the nature of pharmaceutical competition.\textsuperscript{19} As we know, the outcome was the HWA-compromise, where one side (Title I) dealt with post expiration market conditions while the second (Title II) took up patent term restoration for pioneers. The compromise was not immediately successful, however.

In approaching the compromise, generics proposed the Abbreviated New Drug Application (ANDA). This was actually a re-introduction of the pre-1962 situation that would now

\textsuperscript{18}H.R. 3589

\textsuperscript{19}There was lightning-attempt to pass a restoration act alone in 1982, which was narrowly defeated. (H.R. 6444; 1982).
be available for all modern generic applicants. The ANDA would allow generic companies to enter the market more rapidly post-patent expiration and without incurring clinical testing costs. With the mature science supporting the safety of such entry, generics pressed their advantage with the prescription prices paid by the elderly. The elderly are both high-consumers of drugs and often subsist on fixed-incomes (Desrosiers, 1990). At the time, the elderly were only 11% of the US population but consumed 30% of the annual prescribed drugs. Drug expenditures were on the order of 6.5% of total health-care spending in the U.S. in 1983. (Desrosiers, 1990, n. 131, p. 150).

The innovative sector had to roll with these punches as more were to come. A concern over disingenuous use of the patent system by firms was emerging as part of the debate. Rep. Waxman had initially proposed limiting PTR to drugs whose patent terms had expired leading up to approval (Fox & Bennett, 1987, p. 60). That limit could however create an incentive to let one's patents expire in order to earn restoration status. The threat of creating moral hazard, where PTR would incentivize the firm to lengthen clinical trails in a bid for longer effective patent terms, became a new problem for the compromise.

There were several strategic options available to an innovator which would delay patent issuance garner extra years of protection. Chief among them is the practice of filing patent continuations. Continuations make patent prosecution continue forward, preventing issuance and the (then) start of the patent clock (Lemley & Moore, 2004). In broad terms, continuations allow the patent applicant to modify their claims about their invention. There is no legal barrier to modifying one's claims, so long as one does not add new information to cover a competitor's product (Lemley & Moore, 2004, p. 77).

---

20 Chapter 4 discussed the abbreviated pathway which existed prior to the 1962 Kefauver-Harris Amendments but had been removed as part of the focus on safety and efficacy. Prior to 1962, the ANDA pathway required no human or animal testing for generic entrants. Generics merely had to be chemically similar to the branded drug (Desrosiers, 1990, p. 137).

21 Administration of the application process also provides inefficiencies. These include a general backlog at the Patent Office, in addition to general bureaucratic delays. Administrative lag between filing for a patent and receiving a grant can range from only a few months to several years (Grubb, 2005, p. 8).

22 Kingsdown Medical Consultants v. Hollister, 863 F.2d 867 (Fed. Cir. 1988).
Despite the negative treatment in the literature, continuations have positive attributes in helping innovators respond to comments and critiques issued by the patent examiner. Filing a continuations is a formal way for the inventor to update the patent’s written disclosure to match changes in the state of science during application review (Hegde et al., 2007). Naturally, writing a technical document encompassing science and legal nuance makes patent application a difficult process. That process is unlikely to be completely successful in a first attempt, and continuations helped to structure the revision process, supporting the conversation between applicant and the patent examiner.

With a difficult and rapidly evolving science, it is perhaps no surprise that continuations were especially prevalent in the pharmaceutical sector (Lemley & Moore, 2004). Lemley (2004b) notes that the patent continuation procedure, despite its editorial upsides, nearly always becomes detrimental to society. Groups opposed to the tentative HWA’s restoration calculations then called attention to these and other existing strategic “games” innovators could play with the patent system. Gaming the patenting process was seen as any way in which firms were

...“obtaining product, process, use, and other kinds of patents to extend their protection well beyond the original patent period and that they manipulate and delay patent issuances to obtain patent extension through existing patent procedures (Lourie, 1984, p. 535).”

The strategic approach to patenting took on the new comprehensive title of “evergreening” and an anti-evergreening clause found its way into bill revisions. The particular amendment in the HWA debate made a pharmaceutical patent un-extendable if the information in its disclosure was already disclosed or claimed in a previous application. This

---

23 There are also patent reissues (35 U.S.C. §251), were simple drafting errors that damage patent protection are corrected. The newly issued patent was then “corrected”, but could not be enlarged in scope. See Mossoff (p. 1003, 2007), citing Blake v. Stafford, 3 F. Cas. 610, 612 (C.C.D. Conn 1868) (No. 1,504).


would effectively place any PTR onto the earliest of the drug’s relevant patents.26 While it would not directly address the patent system’s existing evergreening incentives, it would prevent any strategic combining of PTR and evergreening.

Despite the diminishing bargaining situation for innovators, the anti-evergreening clause was later weakened.27 The Judiciary Committee and the Patent Office were not of the opinion that evergreening was a large problem (Fox & Bennett, 1987, p. 114). Further, even if reports of evergreening did reveal an emerging ability for firms to approach patenting in a strategic fashion, the moral hazard was not particular to the pharmaceutical industry. The weakening of extraneous patenting limitations helped to maintain other parts of the negotiations and move toward agreement (Engelberg, 1999).28 The evergreening clauses were eventually removed entirely. That nearly met with a conclusion to the bargaining.29

Some concerns over incentivized behavior remained, however. These were included in revisions but remained more focused on the pharmaceutical dynamic specifically, such as adding a restoration cap (American Enterprise Institute, 1981, p. 19-20). The average eroded period at the time was, in fact, seven years (Lourie, 1984, p. 528) so a cap of that length would remain in line with restoration while preventing any extension.30

The debate also hit on the idea of attaching restoration to only the oldest of a drug’s relevant patents.31 While it would limit strategic behavior, it was too coarse and had a

---

26 "For example, if the approved product is the subject of several patents as a result of filing continuation, continuation-in-part, divisional or otherwise related patent applications, each of which discloses the approved product and the approved use, then only the earliest patent is eligible for extension." See House Report on HR3605 Part I, at 37-39.

27 The weakened provisions precluded only future patents that disclosed an identical compound as a previous application for the same use. Patent continuations which had enough difference from an original disclosure could still then extended instead of their earlier versions.

28 I. This is also not to say there are not other, more amorphous incentives to creep one’s patent term forward (Tyson, 1999, p. 206).

29 S. 2892 and H.R. 7952, 96th Congress; see Lourie (p. 527, 1984), AEI (1981)


31 "For example, if the approved product is the subject of several patents as a result of filing continuation, continuation-in-part, divisional or otherwise related patent applications, each of which discloses the approved product and the approved use, then only the earliest patent is eligible for extension." House Report on
large flaw. The limitation could effectively remove extensions for products which the Act was specifically targeting.\footnote{Minority Views of Rep. Bliley, House Report on HR3605 Part I, at 75-76, cf. Fox & Bennett (p. 111, 1987).} We noted earlier that firms often file patents early in research phases to protect even the most embryonic developments in the highly competitive R&D. Broad, early patents, often called "genus patents", may disclose the information in a later, more specific and drug-related "species" patent. That species of use only emerges later in drug development, where the additional utility is patented. As both patents are owned by the same inventor, and thereby within the same firm, the later patent will not infringe the first. Forcing the patent term extension to reference the early genus patent would however preclude any effective extension of the later species patent, which in many cases is the patent which covers the drug proper (Fox & Bennett, 1987, p. 111). That would be antithetical to the goals of the patent system and the inclusion of restoration.

A final compromise was achieved when the sides agreed to let the innovator choose a single patent within the package of patents accompanying a new drug to receive the extension. The single patent could be any of the applicant's choosing and served as the recipient of the extension and also lent its starting date to the calculation of erosion.\footnote{Remarks of Rep. Waxman, Cong. Rec. of Sept. 6, 1984 at H9130 and H9132. See also §156 (c)(4).} The HWA then calculates the restoration refund using the length of the regulatory review period and the issuance date of the chosen patent.

Although larger concerns about gaming the patent system had emerged, weakening the stance of the pharmaceutical industry's claims to need enhanced patent protection, the general claim for restoration remained accepted through the debate. During Congressional Testimony, pharmaceutical firms claimed that they required 12- to 19-years of market profits
in order to recoup R&D expenditures (Desrosiers, 1990, p. 125). Despite some unprotected overlap with the known erosion spans, the figures had general acceptance as a guiding line for restoration. A total cap of five years, rather than seven, was imposed, and the bill moved toward approval.

The big picture that emerges from a history of the legislative process which gave us the HWA is one of minimal attention to the patent term restoration itself as larger concerns about costs of drugs, the remaining barriers to generic entry after patent expiration, and the strategic use of patenting procedures took over. In fact, there was very little consideration of how to calculate the refund itself, other than attaching the arithmetic to the cause of the erosion. We discuss the calculation in the next section. The provisions and administration of the restoration reflect the concern with disingenuous use of the system, and moreover highlight the lack of formal discussion on the patent term restoration rate itself.

5.2.1 The Restoration Calculation

In order to award a patent term restoration, we must establish the length of the regulatory review period. To qualify in the calculation, the regulatory review must, in fact, cause erosion. That is, only time where the patent term overlaps with the clinical trials period may be claimed for extension. In early practice, nearly half of applicants had IND dates that predated their significant patent issues (Lourie, 1989). Patents at that time began on the date of issue whereas they now begin from the date of filing, a fact which makes much of the concern over continuations and keeping patents pending moot today. It bears noting however that there was a rush to file patent applications prior to 1995’s change from the date-of-issue to date-of-filing measurement (Lemley & Moore, 2004, p. 85).  

\[34\text{Patent Term Extension and Pharmaceutical Innovation: Hearing before the Subcommittee on Investigations and Oversight of the Committee on Science and Technology, 97th Cong., 2nd Sess. 18 (1982). That range included a normal return of eight to ten percent per year.}\]

\[35\text{Up to 1989, the average overlap was 2.5 years, with a minimum of 10 days to a maximum of 7.7 years (Lourie, 1989, p. 174).}\]

\[36\text{In light of the discussion above and the vocal concern over the gaming of the patent system, this could be seen as evidence that the extra 3 years of statutory patent length were less valuable than the issuance-}\]
After FDA certifies the approval data, the successful drug candidate files a patent extension application with the USPTO. The application is then forwarded to the FDA with a request for preliminary comments (Tyson, 1999), including confirmation of approval with supporting data from newly completed clinical trials. If the FDA raises no eligibility issues, they implicitly certify that that this compound did lose patent time while conducting clinical trials. The USPTO then asks the FDA to determine the length of the regulatory period. The FDA is solely responsible for this action (Tyson, 1999). Although we have lumped the regulatory period into one general period in the discussion, a period coextensive with the clinical trials period, the FDA actually measures two periods: from the Investigational New Drug Application (IND) filing to the New Drug Application (NDA) submission; and the submission through approval. That latter administrative period is treated differently than the clinical period. It garners a full patent term refund as the erosion there is inflicted solely by the government. The length of the clinical period is however under some degree of control by the firm. We return to this below.

As the IND is a passive process, establishing an exact start to the IND causes minor mistakes with respect to HWA calculations. The receipt date is the correct date, which is naturally difficult for the applicant to determine for inclusion on their initial application.\(^{37}\) The period determined by the FDA’s records is published in the Federal Register. It becomes final after 180 days if it raises no complaints.

The publication is important in that third parties may raise concerns here. In a nod to evergreening and moral hazard issues, competitors are deputized to question the calculation and the impending patent term adjustment prior to official restoration. This prominently-placed provision is however seldom-used. The option is available though for competitors to claim and then show a lack of due diligence on the part of the innovator in pursuing

\(^{37}\)The pathways established for due diligence complaints also allow for error corrections with regard to these start dates.
their clinical trials. Though interesting in theory, the adjustment procedures recorded from the HWA’s enactment until now are almost exclusively initiated by minor differences in the mailing and receipt dates of IND applications (Malkin, 1999, p. 212). This is an administrative protocol only, and is not seen as strategic, amounting to a change of a day or two (Malkin, 1999).\footnote{Calculating the span has other small but specific caveats. They have largely ceased to be applied, having existed for the transition periods into the HWA and over the change in patent term straddling 1995. Also, as of August 21 2007, there is a limit on the number of continuation applications. Though less useful to extending an expiration under the current date-of-application measurement, continuation filings are limited to two (Reed Smith eds, 2008).}

In theory, the PTR refund would also be reduced through any finding of insufficient diligence by the firm. Due diligence is defined as a “degree of attention, continuous directed effort, and timeliness as may reasonably be expected from, and are ordinarily exercised by, a person during a regulatory review period.”\footnote{\textsection 156(d)(3), 98 Stat 1599-60} A party which believes the applicant did not act with due diligence during clinical trials files a petition for a revision of the length of the regulatory review period. The burden of proof lies with the applicant. Although they may be at an information disadvantage here, the petition must only show “sufficient facts to merit an investigation by the Secretary.”\footnote{House Report Part I, p. 41-42, Fox & Bennett (p. 174, 1987).} Sufficient facts would include such coarse statistics as the failure to begin clinical trials for an unreasonably long time after the FDA accepts an IND or trials taking an unreasonable amount of time for the science involved.

Without wading further into this discussion, we may conclude that, at a minimum, the lack use is not due to any lack of access. There are actually two periods that could be used to bring down the calculation.\footnote{21 CFR 60.24. They can similarly file a petition with claims on the length of time within the regulatory review period where marketing approval was not diligently pursued (21 CFR 60.30).} Further, any third party can request a review of that calculated period within 60 days of publication in the Federal Register (Boone, 2009); there is no particular standing requirement here. Despite the bounty, the person who most often files turns out to be the applicant themselves. Even that number is small, however. From 1992 up through 2010, only twelve petitions have been filed.\footnote{75 FR 61494 (2010). Available: https://www.federalregister.gov. “Agency Information Collection A-}
Lourie (p. 356, 1985), (p. 552, 1986) noted at the time, and it has born out in practice, that while a diligence concern was intuitive and perhaps even visceral at a time when disingenuous gaming of the patent system was emerging, it would be hard for the firm to achieve in reality. Difficulties in managing trials at that level of granularity are manifold. Stalling the initiation of trials leaves all the uncertainty and clinical hurdles ahead. Sitting on one’s NDA docket is arguably irrational, given the highly-competitive milieu of pharmaceutical innovation.\textsuperscript{43} Moreover, the costs to those trials have already accumulated and will have put pressure on the firm’s existing revenue streams. Further, the refund is bounded by the 5-year cap. Any incentivized behavior would operate within a rather tight boundary (American Enterprise Institute, 1981, p. 19-20).

The digression on diligence reveals something interesting though: there is no evidence of due diligence problems. Nevertheless, the legislative debate placed a lot of emphasis here, and a number of brakes against incentivized behavior remain in the HWA. There is a 5-year maximum, any third-party may investigate the applicant’s behavior in the clinical trial period and, most boldly, the innovator does not receive full restoration of their clinical erosion. The overlapping period of patent term and clinical trials, between the IND and the NDA filing, is divided in half. While the time between the NDA filing and final FDA marketing approval is restored in full, the larger portion gets one-for-two consideration.\textsuperscript{44}

\textsuperscript{43}There are some conspiracy theories often put forward of drug companies burying or scuttling revolutionary or market unsettling inventions. The narrative is often posited with a competitor firm who has a drug that would benefit from the release of the other. Notably, this comes up in historical notes on military adversaries. See Lesch (p. 84-85, 2006).

\textsuperscript{44}Also notably, there have been few HWA issues debated in court. Admittedly though there are rather
This one-half refund rate is striking in several respects. If the goal of the bill is restoration, bounded already by the existing statutory grant, why is the patent not fully restored? We may hypothesize some reasons, aided by the preceding discussion on moral hazards in the patent system. In trying to answer the question more formally, one is struck by the lack of discussion in the legislative history as to the rate itself. Although the hearings make heavy use of the patent narrative, with both sides acknowledging the necessity of the remunerative role played by patents, the historical view is unclear on what was expected from the refund parameter itself.

What follows is the best available historical explanation for the choice of a refund rate. While we must assume the existence of back-room conversations, the published record and associated literature confirms, at a minimum, a lack of economic consideration in its determination. The formal modeling in Chapter 6 extends to address this shortcoming.

5.3 Historical Explanation of the 50% Refund Rate

The inclusion of anti-evergreening provisions in the history of the HWA was clearly a nod toward preexisting issues with the patent system at large. From that point in the debate forward, conversations circle around the various ways in which firms could be predicted to game the existing system and any restoration provisions. References to the refund rate itself, where they exist at all, make reference to the concerns about evergreening, continuations, and the incentive to aim for the broadest patent claims during the patenting process. Though the major anti-evergreening provisions were excluded from the final bill, pieces of control remained in the form of the one-patent-choice and the 5-year total cap. The most major restriction was, of course, the 50% refund rate.

After nearly three decades of use, the—arguably complete—lack of due diligence com-

few drugs coming off patent at any given moment for whom the HWA is an option. The USPTO therefore handles only 50-60 applications for extensions per year (Tyson, 1999). The total amount of Hatch-Waxman litigation remains very small compared to the work of patenting (Mahn, 1999) and the lack of litigation can only be a small anecdote.
plaints is striking. It may indicate that the refund rate kept the firms diligent. That may also be an artifact of the uncertain and competitive conditions within the pharmaceutical industry. The 5-year cap may also play a role, as could the intervening change in statutory patent terms from the 17- to 20-year systems, with the other changes that entailed. With the exception of the last point, these issues could all have been debated and weighed with the fervor seen on the generics side of the debate. Despite the importance of the refund rate to the whole concept of patent term restoration, it emerged largely as a remainder in the legislative calculus rather than a central parameter.

That is not to say the fractional refund was entirely ad hoc however. The point is rather that its appearance in the HWA was not a result of legislative compromise or detailed analysis. One might be forgiven for calling it a random occurrence. In 1981, H.R. 6444 was sponsored by Rep. Kastenmeier as part of a triumvirate of PTR-legislation.\footnote{H.R. 6444; S.225, the successor to S.2892; and H.R. 1397, which was essentially the same as S.225} This early bill contained a one-for-two (50\%) extension for clinical trial time over ten years; there was a full one-for-one extension for the first ten years of clinical trials.\footnote{Hearings on the PTRA of 1981, at §155 2(A) (proposed), "subject to subparagraph B, the term of the patent shall be extended by the time equal to the regulatory review period for such product or method for the period up to 10 years after the date of filing of the earliest application for the patent and the time equal to one-half the regulatory review period for the period between ten and twenty years from the date of the earliest patent application."} The bill's forerunners had no provisions for fractional extensions or this tiered structure, preferring a simple grant of the full clinical time up to a cap of a seven years.\footnote{Hearings on the PTRA of 1981; H.R. 1937 §155(b)(2) provides restoration "...equal to the regulatory review period...no longer than seven years." S. 255 states the same in §155(a)(1): "...shall be extended by the amount of time equal to the regulatory review period."}

A seven year cap would have made a one-for-two restoration past a 10 year point largely irrelevant in practice.\footnote{The seven year cap is included in H.R. 6444 as well. §155 2(B): "In no event shall the term of any patent be extended more than seven years. No extension of a patent may exceed twenty-seven years from the date of filing of the earliest patent application. If the term that the patent would be extended is less than one year, no extension shall be granted."} Regardless, the legislative history finds all proposals incorporating a restoration component and some maximum cap past this point in time. That is, up until the sudden appearance of the 50\% refund, which seems to emerge out of thin air having
generated no floor debate on the previous full refund bills. Only one author comments on the emergence: Alan Lourie (1985b). Lourie (1985b) locates the one-half in a draft debated under the Patent Term Restoration Act of 1981. These were the bills that contained only provisions on PTR and lacked any generics entry provisions that were to become the major part of the HWA bargain.

Supporting the assertion that the refund rate was not a source of contention at the time, perhaps generating major back-room discussions while not yet emerging to the floor, this PTR-only bill almost made it into law. The act died in the 97th Congress due mainly to circumstances unrelated to their content (Lourie, 1984). Weather, in fact, was a major factor in the voting failure as many members could not return to vote.

After the near-passage, the congressional records from 1983 and 1984 show the open war between PTR and generic interests. During this time the agro-chemical industry watched the wrangling with growing concern. Their interests motivated H.R. 5529, a sibling PTR-bill, which borrowed parts of 1981’s H.R. 6444 as framing but whose main goal was to keep agro-chemicals out of the compromise. They wanted no part of an emerging compromise for post-patent entrants (Lourie, 1985c). In using an old bill to exclude themselves, they were not concerned with the refund rate itself. They copied the part of H.R. 6444 which included the one-for-two refund structure, now for all lengths of trials, and added the restoration limits from the ongoing debate. From this point forward, the one-half refund rate is in all revisions.

The legislative history is remarkable in its silence here. As noted above, there were significant dealings off-the-record and contributed to the lack of written records. The House Report, whose intention it is to clarify meaning and intent of legislation, simply lists the one-half refund rate along with other provisions. The congressional record does not record any debate on the choice of refund rate. Mossinghoff (p.190, 1999), who was the Commissioner of Patents and Trademarks at the time, reflected that “[t]he length of the exclusivity periods are strictly arbitrary legislative numbers pulled out of the air.” Lourie, (p. 175,
1989) takes a more measured stance, explaining that the fractional refund was based “apparently on the theory that a part of the testing would have to have been carried out by any responsible company and would not therefore be attributable to regulatory review.” An ethical pharmaceutical company would have to spend time between patenting and market entry even if the FDA did not review their safety data.\textsuperscript{49} Some patent term would then be eroded and arguably to the same extent that non-pharmaceutical innovators expend during their non-marketed, patented period.

Clearly, one might start an argument from that point and build toward the fractional rate. But it does not explain the 50% relative to 33%, 75%, or a variable percentage based on other criteria. The lack of consideration here is strange considering the lengthy discussions on aligning incentives which ran through the due diligence and evergreening debate.\textsuperscript{50} After all, a fractional refund rate should put limiting pressure on both due diligence and evergreening but may also fail to incentivize the projects for whom erosion was a significant enough worry to spark the whole PTR conversation.

Speculation aside, the ambiguity, combined with the many caveats and “buts” in the legislation has contributed to the view that it is, as a whole, “inelegant” and “overly complex” legislation (Engelberg, 1999, p. 391-392). In particular, U.S. Supreme Court Justice Scalia stated “[n]o interpretation we have been able to imagine can transform §271(e)(1) into an elegant piece of statutory draftsmanship.”\textsuperscript{51} He drove the point home also, claiming that the language is “not plainly comprehensible on any’s view.”\textsuperscript{52}

This all leads the economist, steeped in the literature on optimizing patent terms (Chapter 2), the regulatory hurdles of drug approval (Chapter 4), and the industrialized science which straddles these realms (Chapter 3) to question the efficiency of the HWA. Is the current favorable view a fortunate outcome from a shot-in-the-dark policy or is there some-

\textsuperscript{49}This sentiment was initiated by Rep. Waxman in his stance against H.R. 6444 in \textit{Hearings on the PTRA} of 1981.


\textsuperscript{52}\textit{Eli Lily & Co. v. Medtronic, Inc.} at 669.
thing to the bargain that approximates a theoretical efficiency? Early studies of the HWA
certainly had their doubts as to the optimality (H. Grabowski & Vernon, 1986). The debate
has not let up since and was only rekindled with the debate over biologic and biosimilar
regulation.\textsuperscript{53}

The final chapter takes up the unaddressed question of the efficiency of the implement-
tation of patent term restoration as an ex ante incentive within the patent narrative. In
addition to the perspective granted from the preceding chapter's overviews, our model takes
up the real-world concerns brought forth by the actors involved in the legislative process.
As this chapter showed, they had a demonstrable role in the form of PTR that emerged
as the Hatch-Waxman Act of 1984. What remains is to determine what the lack of formal
economic considerations in the legislative conversation means for the social efficiency of the
policy.

\textsuperscript{53}See especially the spirited discussion between Kotlikoff (2008), Brill (2008), and Grabowski & DiMasi
(2009).
Chapter 6

Incentives and Welfare Effects of Patent Term Restoration

Their efforts, even when honest, seldom accomplish any appreciable good.

H.L. Mencken

This chapter considers a model of pharmaceutical innovation which proceeds sequentially through the two stages of R&D: preclinical research and clinical (human) trials. Patents play a key role in incentivizing ex ante R&D investment in an uncertain project. The promised duration of a patent, together with the profit flows projected for the project’s market, determines whether the firm will break-even on their investment. Without the patent, competitive entrants may free-ride on the project’s success and reduce or eliminate the profit flows to the inventor. Projects with an expectation of generating ex post revenues sufficient to overcome ex ante R&D costs are therefore incentivized by securing the profit flow for the inventor.

For the pharmaceuticals, we include a patent term restoration (PTR) policy that appends time to the end of the statutory patent grant. PTR was implemented for successful new drug introductions under the Hatch-Waxman Act (HWA) of 1984. The policy restores
a portion of patent term lost to the second stage of R&D, the clinical trials required to secure marketing approval from a regulator. Because the restoration is a function of time lost during clinical trials, the patent-plus-PTR system adds another consideration for the firm's investment in the second stage of R&D. The patent's expanded role now includes incentives for the timing of entry, determined via the firm's choices over clinical trials. The model therefore captures the combined incentive of a static patent term plus a restoration period on pharmaceutical R&D decisions.

The expanded policy goal remains to maximize social welfare. Both systems have the same coarse effects on welfare: less generous policy allows earlier competitive entry of substitutes at the end of the (restored) patent term while more generous policy expands the possibility of a project breaking even. Although the latter benefit may lead to more invention and a higher level of aggregate social welfare, the focus of this essay is the incentives for a single pharmaceutical project. By modeling the ex ante investment incentives we show how PTR affects the firm's perspective on completing clinical trials and the resulting effects on social welfare through changes in drug introduction- and patent expiration-times.

6.1 Two Stages of R&D

Pharmaceutical R&D proceeds through two general stages: preclinical laboratory research and clinical trials. The dividing line is the involvement of human subjects in the clinical stage; preclinical work is conducted largely in a laboratory, proceeding up through animal studies. In the research stage, the pharmaceutical focuses on the discovery and screening of promising molecules. Promising molecules are those known to have some pharmacological activity at a target site, where target sites are parts of a biological process in which deviations can lead to illness. While many discoveries have measurable pharmacological properties, the innovator spends time and effort to screen them for the most promising profiles visible from non-human experiments. A discovery which has enough preclinical data
to suggest a safe and efficacious use in humans may then proceed to clinical trials.

The clinical trials stage is required by the regulator—the FDA in the US—to rigorously support the claims made by the pharmaceutical firm when marketing the drug. Regulatory approval is, in fact, the regulator’s acceptance of a marketing application’s claims that a therapy is both safe and effective when used for the stated indications.\(^1\) The indications claimed by the manufacturer and substantiated through the trials therefore define the market that the drug enters. Discoveries whose preclinical profiles define a sufficient market size to remunerate the already sunk costs of preclinical work and the upcoming costs of clinical trials are the products which eventually make their way into medicine.

This essay focuses on the small subset of all R&D discoveries which have a sufficient safety and efficacy profile and a market profile under the existing patent protection scheme which can justify completing clinical work on top of the research expenditures. We define this subset as containing *marketable discoveries* for clarity. Potential market size is, to a large degree, determined by the incidence of a disease and the ways in which a human may manifest a disease profile. The fact that disease incidence is often over-determined means that a drug may not work for all sufferers. These are scientific factors which are treated here as parameters; the underlying state of science is stable and fixed at the time in which the pharmaceutical firm decides on R&D investment. With a stable science and preclinical work completed, the firm may project a market profile and judge whether patent policy is sufficient to justify further investment and, more importantly, determine how to proceed into human trials.\(^2\)

\(^1\)21 CFR §312.3(b); §314.50, 2012). There is a gray area most notably occupied by dietary supplements and herbs. These molecules are marketable without clinical testing but limited to supporting body functions. They must carry the following warning when claiming such a supporting function: “This statement has not been evaluated by the Food and Drug Administration. This product is not intended to diagnose, treat, cure, or prevent any disease.” 21 CFR §101.93(a)(3)(c) (2012).

\(^2\)This essay considers patents as the sole provider of exclusivity and therefore of market profit opportunity. Pharmaceuticals also have regulator-provided protections which are wholly separate from the intellectual contribution of the drug discovery and may extend profit opportunities past patent expiration. This was less of a concern for small-molecule medicines but is more substantial for newer biologic endeavors. For an introductory discussion on the overlap of protections vis-à-vis optimizing innovation policy, see Eisenberg (2001).
Although the term of a patent is fixed, pharmaceuticals enjoy the unique provision of patent term restoration (PTR). PTR refunds a portion of any patent term lost due to conducting clinical trials. Due to competition in research, pharmaceutical firms generally patent discoveries early in the R&D process. Therefore, their products have a shortened effective patent term owing to the overlap of the statutory patent term and the clinical trial period where the drug cannot be marketed. Shorter trials leave more effective term remaining to accumulate profits while longer clinical processes leave less time. For the US, PTR follows the Hatch-Waxman Act, which refunds one-half of the patent time lost. The particular rate was not the subject of much debate at the time of enactment, but the Act itself is nevertheless regarded as a success (Frank, 2007).\textsuperscript{3} By understanding how PTR policy changes firm decisions towards clinical trials and building upward, we may reach the larger policy questions of increasing welfare.

6.1.1 The Clinical Period

With preclinical work completed and expenditure $k$ sunk, the innovator understands enough about a potential new therapy to estimate the prospective market. The market profile, or the graph of monopoly profit receipts over time $\phi(t)$, is outside the influence of the firm. It owes its description to the scientific and technological work conducted in the preclinical stage.\textsuperscript{4} Generally, the market profile grows from approval onward as doctors and patients accept the new therapy (H. Grabowski, 2008).\textsuperscript{5} More specifically, we conceive of an invention

\textsuperscript{3} The success of the HWA is often weighed on balance between its two major facets. The Act created both the PTR extension for innovative firms and lowered the entry barriers to generic competitors after the restored patent expired. The latter dynamic enjoys a deeper literature than the former, as a generic drug industry was largely absent and unable to provide competition post patent expiration prior to the HWA (H. Grabowski & Vernon, 1996).

\textsuperscript{4} Market size is known to be a significant driver in the further development of potential pharmaceuticals (Acemoglu & Linn, 2004). The current model can extend to include an incentive to conduct directed research, aimed at expanding the emerging market definition, as a separate incentive from pure discovery or drug screening.

\textsuperscript{5} Grabowski (2008) finds empirical evidence showing, on average, a decrease in profit flows towards the end of the the patent term. This is suggested to be a function of market saturation, the entry of non-infringing substitutes, or the withdrawal of marketing resources allotted to the therapeutic, adjusting marketing resources to other therapeutics in the firm's portfolio as the other patent clock runs out.
as generating a new flow of social surplus $\omega(t)$. The firm's profit profile $\phi(t)$ is the profit extracted from the total surplus at time $t$ and therefore reflects the firm's ability to—or choices to—profit on $\omega(t)$.\textsuperscript{6} Here, the firm follows the monopolist's single-market pricing decision to maximize the profits at time $t$.\textsuperscript{7} The magnitude and the divisions between profits $\phi$, consumer surplus $m$, and deadweight loss, $l$ are shown in Figure 6.1.1.\textsuperscript{8} The figure also introduces the assumption that marginal production costs for a drug are negligible relative to R&D costs.\textsuperscript{9}

The patent on a potential drug starts on the same date as the clinical trials. This is because new drugs, being on average protected by more than one patent, must chose a single patent to provide the dates for, and serve as the recipient of, PTR under the HWA. Because the firm discloses information to the regulator in order to begin human trials, under existing patent law they start the countdown to claim rights to that information (Hope, 2008, p. 56).\textsuperscript{10} While an innovator may have several patents over the therapeutic running from a date prior to clinical trials, connecting trials and the PTR-patent term by a common start date provides a realistic minimum for the effective patent life.

The pharmaceutical firm is seen as capable of designing multiple options for performing...

\textsuperscript{6}Although patent protection covers the therapeutic invention's utility, it does not necessarily convey full monopoly power. Owing to the over-determined nature of disease, a second therapeutic molecule may address the same indications in a manner sufficiently different so as to avoid infringing on first patent rights. Such competition is legal and permitted as the regulator does not base approval on patent claims except for situations of generic entrants who are in effect using the innovator's clinical data to support their own entry market.

\textsuperscript{7}Pricing decisions are particularly complicated by the reimbursement policies and bargaining situations of third-party payers, including government programs. The view of $\phi(t)$ as the monopolist's simple profit is however a lower boundary to profits, which is in line the goal of defining minimum sufficient patent policy.

\textsuperscript{8}A decrease in the profit flow after a certain point in the therapeutic's patent's life, as illustrated in the average profile shape found by Grabowski (2008), would be reflective of the firm's decisions to use a different profit strategy over $\omega(t)$, rather than a profit contraction based on a demand shift.

\textsuperscript{9}While marginal production costs for small-molecule drugs are generally small relative to R&D expenditures, next-generation biologic medicines may deviate from this standard assumption (Trusheim, Aitken, & Berndt, 2010; DiMasi & Grabowski, 2007a).

\textsuperscript{10}35 U.S.C. 102 et seq. Outside of legal nuances, there is a competitive concern that a competitor will move to patent their work in a similar research space during preclinical stages. Holding all information about preclinical work secret helps prevent strategic, offensive patenting. While a firm might claim their patent rights within a year from revealing information about their project publicly—which includes revelations to the FDA where there might otherwise be an expectation of privacy—any leaked information may spark offensive patenting as competitors move to take advantage of the information on another's R&D directions.
the clinical trials that will satisfy prevailing regulatory hurdles.\textsuperscript{11} Naturally, a firm does not choose $\tau$ directly but the length—and equivalently, the entry date—is the realization of the many choices which form the pathway to approval. Although approval is never certain, the clinical process is modeled as deterministic by leveraging the definition of a marketable preclinical discovery. Under this definition, the firm does not leave preclinical trials until their level of understanding supports the formation of a hypothesis on safety and efficacy. The hypothesis’ claims are then supported through human trials.\textsuperscript{12} The distance separating the model’s assumption of clinical certainty and the reality of late-stage clinical drug failures is the emergence of new information about the pharmacological

\textsuperscript{11}For discussions on modern trial designs, benefits, and drawbacks, see, i.a., Friedman (2008) and Cato et al (2002).

\textsuperscript{12}Note that under this definition, the separation of pre- and clinical stages could be drawn at a later point—such as the junction of Phase II and Phase III clinical trials—by reinterpreting the clinical choice variable slightly. A firm at the Phase III junction could be seen as choosing a remaining $\tau$, which is added to the effective patent time already sunk into Phase I and II clinical work.
properties of the therapeutic during clinical trials. The modeling outcomes are therefore
bounded but remain aligned towards describing incentives within a minimally sufficient
patent policy. Additional uncertainty will only increase the requirements on the profiting-
period to remunerate R&D investment.\footnote{Changes in information sets stemming from preclinical and clinical work, and the impact that has on clinical choices, is the subject of ongoing research following the fundamentals presented in this model.}

There are sufficient choices within the space of trial designs such that $\tau \in \mathbb{R}$ and $c = c(\tau_i)$ is a $C^2$ function. The innovator contemplating clinical trials design searches the set of designs
which provide a realistic time for meeting the regulator's questions, and which will cover the
sunk $k_i$ costs, while leaving enough patent life to satisfy the break-even criteria.\footnote{Formally, the feasible set of clinical trials options for a given PTR policy $\alpha$, the stated market profile $\phi(t)$, and the projected clinical costs for completion $c(\tau)$ is $D(\alpha) = \{\tau_i\} \left[ \int_{L+\alpha}^{\tau} \phi(t) \delta(t) \, dt \geq c(\tau) + k \right]$.}

Projects must break-even during the effective patent life because competitive conditions are assumed
to obtain after patent expiration.\footnote{Though standard to assume patent expiration allows sufficient entry for competitive outcomes, significant generic entry does not always prevail in pharmaceuticals (Regan, 2008; Kong & Seldon, 2004).} The choice of a trials design then specifies the market
entry date and, through the restored-patent policy, the patent's expiration date.

The profit profile is itself unaffected by the entry date. The profile describes the market
growth from any entry, $\phi(\tau) = \phi_0$ through patent expiration, which in the HWA specification
occurs at $L + \alpha \tau$. Here, $\alpha$ is the PTR policy parameter and $\tau$ is the length of clinical trials
and therefore the market entry date of the drug. Although nominal profits are equal at
the same relative distance from an entry date, their discounted values differ. The discount
process $\delta(t; \tau)$ begins at $t = 0$, when the firm procures clinical funding $c$. We assume a
general discounting process with a discount rate $r$.\footnote{Notation is suppressed in the analysis.} To align times and values we can
specify profit flow at time $t - \tau$ so that the expression for the discounted market profit
becomes $\phi(t - \tau) \delta(t)$. The present value of the profit profile at an entry date $\tau$, assumed to
be suitable for the break-even criteria and to the regulator, is given by (6.1.1).

\begin{equation}
V(\tau; \alpha) = \int_{\tau}^{L+\alpha \tau} \phi(t - \tau) \delta(t) \, dt - c(\tau)
\end{equation} (6.1.1)
Figure 6.1.2 illustrates the profit profile $\phi(t - \tau)\delta(t)$ for two entry dates, $\tau_1$ & $\tau_2$, to highlight the impact of the choice of clinical trials and entry date. The figure introduces the primary assumptions as to the interaction of $\phi(t - \tau)$, $\delta(t)$, and entry. Delayed entry lowers the entire market profile while shifting it to the right by the amount of delay. The concavity of $\phi(t - \tau)\delta(t)$ is assumed to hold over the effective patent interval, $[\tau_i, L + \alpha\tau_i]$ for entry at $\tau_i$.

![Figure 6.1.2: Nominal and Discounted Market Profits at time $t > \tau_i$ for two entry choices](image)

A concern emerged during the HWA’s legislative debates that PTR would insure against patent term erosion enough to make the later-and-lower market profiles sufficiently appealing. The potential moral hazard, described in the HWA as a concern over a firm’s incentives for due diligence in pursuing clinical work\textsuperscript{17}, would emerge when the profits under the area extended to the right by the lengthened patent outweighs the explicit and implicit costs to

\textsuperscript{17}35 USC §156(c)(1). Due diligence complaints follow 21 CFR 60.24(a) and 60.30; Section 60.40 may be employed for an informal hearing to reconsider determinations under the preceding clauses.
delaying entry through longer trials. Moral hazard would, of course, extend the social losses associated with the market power grant unnecessarily when a shorter patent policy would still justify the R&D project. No evidence of problematic moral hazard exists in the history of the legislation or has emerged in the nearly three decades since the HWA’s enactment.\footnote{The outcome highlights the lack of concern in the debates on the bill from nonpartisan analysts relative to the legislators. See, esp. Lourie (p. 365, 1985). Though administrative avenues exist to bring diligence complaints to the regulator and carry penalties against firms, very few have been submitted to the FDA. (Federal Register 76:8 (January 12, 2011) p.2127). The few that exist focus on minor date changes.} Nevertheless, the concern motivated a limited PTR refund in the HWA itself.

Figure 6.1.2 makes clear that the discounting of future receipts places constant pressure on the firm’s decisions, independent of any refund rate.\footnote{Note there would be competitive advantages to earlier entry when considering competitive R&D games between firms (DiMasi & Paquette, 2004; DiMasi & Faden, 2011).} The figure assumes that at some date the marginal impact of the discount process dominates the growth process of the market. For the illustrated relationship, there is for each entry date a maximum discounted return during the effective patent life. If the effective patent term spans this maximum point we hold that the derivative of the market profile with respect to time is negative at times greater than the maximum. That is, we assume that $\delta(t)\phi_0 + \phi(t)\delta(t) > 0$ $\Rightarrow \phi(L + \alpha\tau - \tau)\delta(L + \alpha\tau) + \delta(L + \alpha\tau)\phi(L + \alpha\tau - \tau) < 0.$\footnote{Subscripts are used to denote partial derivatives with respect to variables and variable parameters. The prime notation is reserved for the derivative with respect to (absolute) time $t$.} While it is possible to relax this assumption, un-dominated market growth could lead to boundary solutions for the maximization; the focus here is on the unconstrained behavior of both the firm and policy.

6.1.2 Research Incentives

Research expenditures in the first stage of pharmaceutical R&D includes laboratory and preclinical—no human subjects—efforts. Again, the goal is to identify and characterize the human potential for a therapeutic molecule. This essay follows the literature and models the discovery phase as a Poisson process (Dasgupta & Stiglitz, 1980).\footnote{Also see, O’Donoghue et al. (2004), O’Donoghue et al (1998a); and Dixit (1988).} The distribution of time intervals separating discoveries is then distributed exponentially. Research investment
$k$ affects the distribution through the Poisson rate parameter.\textsuperscript{22} Larger investments in research discover molecules and reveal their characteristics more quickly, helping to define the target population and the safety and efficacy claims thereon.

The probability that a marketable discovery emerges during an exogenously determined research period $\bar{\tau}$ is $Pr[X < \bar{\tau}]$. As the research period is assumed static, we suppress this time interval by defining a basic probability measure $p := p(k; \bar{\tau}) = 1 - e^{-k\bar{\tau}}$. The function $p$, based on the underlying distribution of discovery times, reflects the the typical characteristics of diminishing marginal returns to research capital: $p_k > 0$, $p_{kk} < 0$. A successful discovery earns the value of the payoff function $V(\tau; \alpha)$ (6.1.1). The firm need not realize the capital expenditure as a loss in the no-discovery outcome under the presumption that any research expenditure goes forward into another research period. With a Poisson discovery process, the probability of discovery in subsequent rounds is again given by the exponential distribution. Further, the process has no memory of the success or failure in the previous term. The question facing the firm is the same as before, making the question in the first stage of R&D stable.\textsuperscript{23}

Incorporating the probability of discovery, we may consider the expected present value of an R&D investment as $Pr[X < \bar{\tau}]V(\tau; \alpha) + (1 - Pr[X < \bar{\tau}])0) - k$. This lends the

\textsuperscript{22}While we reduce notation by assuming the innovator has a research period, $\bar{\tau}$, which is a time horizon which they hope to segue successfully from discovery to clinical trials, for $k > 0$ the Poisson process lends a distribution of intervals between discoveries described by $F = 1 - e^{-ks}$.

\textsuperscript{23}Again, the model can be expanded to account for informational progression through research, potentially raising the probability of success in subsequent rounds for every level of research effort. Longer preclinical work may also have a benefit in the clinical phase, with early work saving on clinical trials design by, broadly speaking, lowering the informational hurdles to approval (Roberts Jr. et al., 2003; DiMasi, 2001). Also notable, recent lattice programming results have generalized comparative statics analysis of optimal stopping problems of which R&D processes are a genre (Quah & Strulovici, 2009; Athey, 2002). A break-even constraint, operating as type of budget constraint that defines the feasible choice set, often complicates these techniques (Antoniadou, 2007). With the addition of a PTR policy that changes the budget constraint as one spends more in the clinical trials phase, the generality of lattice programming is not direction applicable. The generality of that literature's results meshes admirable well though with the need in innovation policy research to design a one-size-fits-all patent system to cover the broadest range of innovation incentivization. The extension of lattice-supported, monotone comparative statics approaches to optimal stopping problems for innovation policy, and especially PTR policy, is advisable.
specification of the firm's complete R&D objective function as

$$\arg\max_{k,\tau} U(k, \tau; \alpha) = p(k) \left[ \int_{\tau}^{L+\alpha\tau} \phi(t-\tau) \delta(t) \, dt - c(\tau) \right] - k \quad (6.1.2)$$

subject to \( V(\tau; \alpha) - k \geq 0, \ k \geq k_{\text{min}}, \ \tau \geq \tau_{\text{min}} \)

Though minimized in its impact in the interior analysis, where \( k > k_{\text{min}}, \ \tau > \tau_{\text{min}}, \) and \( V(\tau; \alpha) > k, \) research expenditure is important in fitting PTR analysis into the overarching patent literature. An unconstrained maximum however leaves the policy maker free in their choice of policy and so gives the broadest perspective on the potential for welfare maximization through PTR.

### 6.2 Optimization and Reactions

Under the preceding definitions, an interior optimum \((k^*, \tau^*)\) is defined for a given \( \alpha \) by the first order conditions. Assuming that \( p(k) \neq 0, \) the optimum values solve:

$$p_k(k) [V(\tau; \alpha)] - 1 = 0 \quad (6.2.1)$$

$$\alpha [\delta(t) \phi(t-\tau)]_{L+\alpha\tau} - [\delta(t) \phi_0]_{\tau} + h(\tau) - c_{\tau}(\tau) = 0 \quad (6.2.2)$$

and \( h(\tau) = \int_{\tau}^{L+\alpha\tau} (-\phi'(t-\tau)) \delta(t) \, dt \)

Rearranging (6.2.2) gives \( p_k(k) V(\tau; \alpha) = 1 \) and the straightforward interpretation of balancing marginal (unitary) costs of additional preclinical research with the marginal increase in the expected value of the profit stream. The role of \( k \) is absent in 6.2.2 under the assumption that the entry date does leave the constraints slack. Rearranging (6.2.2) illustrates the firm balancing the marginal benefits of extending the effective patent term at the expiration date \( \alpha [\delta(t) \phi(t-\tau)]_{L+\alpha\tau} \) with explicit marginal costs \( c_{\tau}(\tau) > 0 \) plus marginal value of the entry day’s receipts, \([\delta(t) \phi_0]_{\tau} > 0\). The function \( h(\tau) \) captures the crescent of
profits lost along the curve between marginally-shifted profiles. The marginal change in \( \phi \) is now negative because the shifting enters through \( \tau \). Moving the negative factor outside the integral reveals \( h(\tau) \) as a marginal cost to longer trials.

Where this interior maximum exists and all constraint conditions remain non-binding, we obtain the firm’s equilibrium reactions to a change in the PTR policy \( \alpha \) by differentiating the implicitly-defined \( \tau^*(\alpha) \) and \( k^*(\alpha) \) in Equations 6.2.1 & 6.2.2. Holding the maximum and boundary conditions and rearranging yields

\[
\frac{\partial k^*}{\partial \alpha} = \frac{-U_{kk}}{U_{kk}} \bigg|_{k^*, \tau^*} = \frac{-\tau^* \left[ \delta(t) \phi(t - \tau^*) \right]_{L+\alpha \tau^*}}{p_{kk} \left( V(\tau^*; \alpha) \right)^2}, \quad (6.2.3)
\]

\[
\frac{\partial \tau^*}{\partial \alpha} = p(k^*)v_{\tau \alpha}(\tau; \alpha) \frac{-U_{kk}}{\det[H]} \bigg|_{k^*, \tau^*}, \quad (6.2.4)
\]

We define \( [H] \) as the Hessian matrix of second-order partial derivatives of Equation 6.1.2 and \( v(\tau; \alpha) \) as the market profile portion of \( V(\tau; \alpha) \): \( \int_{\tau}^{L+\alpha \tau} \phi(t - \tau) \delta(t) \, dt \).

With diminishing marginal returns to research capital, \( p_{kk} < 0 \) implies \( \frac{\partial k^*}{\partial \alpha} > 0 \) at the interior maximum. The slack budget constraint assures \( V(\tau^*; \alpha) > 0 \), which is then squared as well. With a two-stage model, there are no cross-partial effect between \( k \) and \( \tau \) and the concavity of \( U \) rests on \( U_{kk} < 0 \) and \( U_{\tau \tau} < 0 \). Where that obtains, \( p(k) \frac{U_{kk}}{\det[H]} > 0 \). The reaction of the innovator to a change in policy with respect to clinical trials choices, \( \frac{\partial \tau^*}{\partial \alpha} \), depends only on the sign of the cross-partial effect of \( \tau \) and \( \alpha \) on the present value of the market profile. Specifically,

\[
v_{\tau \alpha}(\tau; \alpha) = \left[ \delta(t) \phi(t - \tau^*) + \alpha \tau^* \delta'(t) \phi(t - \tau^*) - \tau^* \delta'(t - \tau) \delta(t) + \alpha \tau^* \delta'(t - \tau) \delta(t) \right]_{L+\alpha \tau^*} \quad (6.2.5)
\]
6.2.1 Social Welfare and Policy

The inclination of the firm’s input choices to policy changes allows the policy maker to optimize social welfare. Policy considers the entire market surplus created by an innovation. This implies a consideration of two periods: the patent period with $\omega(t)$ and the post expiration period, where the total surplus $\bar{\omega}$ is available to society. From Figure 6.1.1, the relationship $\omega(t) = \phi(t) + m(t)$ holds and $\bar{\omega}$ is understood to be the total market surplus at $\phi_{\text{max}}$.

The policy objective is analogous to the profit maximization, including the reframing of $\omega$ to be entry-independent, $\omega(t - \tau)$, because no surplus exists prior to entry.

$$\max_{\alpha} W(\alpha) = p(k) \left[ \int_{\tau}^{L+\alpha \tau} \omega(t - \tau) \delta(t) \, dt + \int_{L+\alpha \tau}^{\infty} \bar{\omega} \delta(t) \, dt - c \right] - k \quad (6.2.6)$$

Welfare maximization is subject to the constraints which dictate the firm’s choice also. Total welfare is maximized where $\frac{\partial}{\partial \alpha} W(\alpha) = 0$ and the firm’s constraints are satisfied. Thus, the optimal $\alpha$ satisfies the first order condition, balancing marginal social benefits to PTR with marginal social costs where the following holds:

$$p_k(k) \frac{\partial k^*}{\partial \alpha} \left[ \int_{\tau}^{L+\alpha \tau} \omega(t - \tau) \delta(t) \, dt + \int_{L+\alpha \tau}^{\infty} \bar{\omega} \delta(t) \, dt \right] = (6.2.7)$$
$$+ \alpha p(k) \frac{\partial \tau^*}{\partial \alpha} \omega(t - \tau) \delta(t)|_{L+\alpha \tau} = (6.2.8)$$
$$+ p(k) \frac{\partial \tau^*}{\partial \alpha} \left[ \omega(t - \tau) \delta(t)|_{\tau} + \alpha \bar{\omega} \delta(t)|_{L+\alpha \tau} \right] + p(k) \frac{\partial \tau^*}{\partial \alpha} c(\tau) \quad (6.2.9)$$
$$\frac{\partial k^*}{\partial \alpha} (1 + p_k(k)c(\tau)) \quad (6.2.10)$$
$$+ \tau p(k) [\bar{\omega} \delta(t) - \omega(t - \tau) \delta(t)]_{L+\alpha \tau} \quad (6.2.11)$$

This expression of the social tradeoff assumes the moral hazard situation, $\frac{\partial \tau^*}{\partial \alpha} > 0$. Without additional assumptions, it is also possible that the opposite relationship holds. We discuss the implications in the next section as they pertain to social policy.
6.3 Discussion

While the generality of the model does not lend itself to analytical solutions it does capture the primary characteristics of pharmaceutical R&D investments relevant to innovation policy inquiry. The major assumptions are broad but capture a wide swath of pharmaceutical R&D project characteristics. They also deliver the expectation of characterizable solutions, if not solutions which lend to general policy recommendations.

The expression of social tradeoffs, captured by the terms 6.2.7-6.2.11, depends on the sign of $\frac{\partial \tau^*}{\partial \alpha}$. The balance above assumes the moral hazard situation, $\frac{\partial \tau^*}{\partial \alpha} > 0$. There, policy perceives benefits to increasing PTR of (6.2.7) & (6.2.8). The former is the marginal increase in the expected value of the monopoly period’s surplus and post-expiration surplus. The second benefit (6.2.8) is the area under the welfare curve to the right of the previous patent expiration opened up by the longer policy. Of course, that factor has an offsetting component as part of the bracketed term in (6.2.9). There, $\omega(t - \tau)\delta(t)|_\tau$ measures the profit interval forgone between entering now and entering later. That term also captures the cost of the delayed start to the maximum surplus period, $\bar{\omega}$. It also adds the explicit additional costs to longer trials, adjusted by the expectation that we get to the trials period. Finally, there is the explicit cost to growing $k^*$ found in 6.2.10, along with the counterintuitive problem of $\frac{\partial k^*}{\partial \alpha} > 0$ raising the probability of spending $c(\tau)$ after the more likely completion of the preclinical phase.

As for the detriment of the deadweight loss experienced in the extended period, (6.2.8) acts as the balance to (6.2.8). Society does have a new sliver of $\omega(t - \tau)$ to enjoy at the new expiration time, but they forgo the benefits of moving to $\bar{\omega}$ and miss out on the discounted value of $\bar{\omega} - \omega(t - \tau)$ at $L + \alpha \tau$.

If parameters instead create the conditions for $\frac{\partial \tau^*}{\partial \alpha} < 0$, terms (6.2.8) and (6.2.9) will switch sides in the equation. While more assumptions as to functional forms and magnitudes could categorize the outcomes over $\frac{\partial \tau^*}{\partial \alpha} < 0$, the fact that the maximizing condition has
oscillating terms may pose problems for general solutions.

Note however that the bounded market assumption plays a central and encompassing role in restraining the chances for moral hazard. It is a rather benign assumption also. In addition to recent work finding this condition empirically\textsuperscript{24} we have limitations provided by regulatory provisions that restrict a drug's use to a list of stated indications. An assumption of a bounded market has multiple supports in the narrative.

Relaxing the market assumptions, or equivalently including regions of rapid market growth inside the effective patent term, still must contend with the discounting process and the revenue-less deserts between R&D and market approval. Owing in part to the lags but also to the large levels of uncertainty, recent empirical work found sector-specific costs of capital to be significantly higher for pharmaceuticals than for other R&D sectors (Harrington, 2012).\textsuperscript{25} The HWA's limited refund parameter of \( \alpha = 1/2 \) certainly doesn't lend momentum in that direction, but the lack of due diligence complaints, does lend support that there is realistically no room for moral hazard.

That leads back to the first-order condition for social welfare maximization. If we may leverage empirical evidence for support that functional forms and magnitudes do not support \( \frac{\partial r^*}{\partial \alpha} > 0 \), then we must exchange the necessary terms to establish the social balance. The number of interactions in the maximization conditions do not readily incline one to judge whether policy is too low or too high even with an assumed comparative static result. For both the single R&D project and for aggregate innovation policy issues, empirical data is needed to support specific choices of functional forms that could support deeper interpretation.

Current events do motivate continued research on this path. There remain many welfare-focused concerns on the relationships between pharmaceuticals and patents including "gaming" and "evergreening" of patents via loopholes in patent law (Bulow, 2004); exploiting

\textsuperscript{24}See especially Grabowski (2008) and DiMasi et al. (2003).

\textsuperscript{25}Also note the energetic debate between Kotlikoff (2009,2008); Grabowski and DiMasi (2009); and Brill (2008).
the requirements of listing patents on drugs in the FDA's Orange Book (Derzko, 2005); and exploiting the generic drug approval processes as a barrier to entry (Kong, 2009). There is undoubtably an incentive to extend market power once the uncertainty of R&D has been resolved. While it is straightforward to attribute these activities to rent seeking incentives, it is equally interesting to question whether more generous innovation policy could reduce negative incentives.

Recent legislative changes may lend a way to answer that question. In the competitive world of medical R&D, a pharmaceutical firm currently profiting from an approved product always faces competition from improvements. For the history of the industry, improvements have generally displaced pioneers at a relatively slow pace. With the development of biotechnology however, the ability to experiment, control, and ultimately innovate at the genetic level is speeding up progress. Developments must still filter through the clinical proving grounds, but the landscape there is changing as well. In updating the way the FDA regulates biologic medicines and, more importantly, by creating an approval pathway for biosimilar drugs—the biologic analogue to generic drugs—we have changed major features in the landscape of R&D competition (DiMasi & Faden, 2011).²⁶

The established bio-pharmaceutical industry admitted a concern over increasing competitive pressures and pressed for additional protections of their profit profiles in the new legislation. Biologic innovators will now have twelve years of exclusivity covering their biologic products after approval, a privilege secured by the FDA as a gatekeeper (Carver, Elkan, & Lietzhan, 2010). This market protection overlaps with the intellectual property protections of patents and the HWA's restorations. With the reinforcement that exclusivity provides for profits, the marginal value of extending the patent term after approval changes relative to other profit maximizing strategies. By strengthening innovation policy through an additional program, there is less pressure on the patent to secure ex post remuneration.

²⁶Also see Vernon et al. (2006) for considerations on how genetic knowledge may change trial design, allowing technology to substitute for clinical trials components.
The new regulatory scheme stands to free the patent system from performing double-duty, incentivizing ex ante investment and then shifting stance to defend a market profile.\textsuperscript{27} Granted, there may well be new inefficiencies in shouldering the FDA with the responsibility of ordering markets,\textsuperscript{28} and fundamental questions remain as to the patent system’s net effect on cumulative innovation in the genetic arena.\textsuperscript{29} PTR will remain however the unique exception to the one-size-fits-all patent regime. With its novel ability to affect choices on clinical development, it warrants renewed attention in the changing landscape.

\textsuperscript{27}The conflation of ex ante and ex post roles for patents has been cogently criticized by Lemley (2004a).
\textsuperscript{28}Although they already do perform this duty in a limited fashion with Orphan Drugs. For a discussion, see Rogovski (2006).
\textsuperscript{29}See Eisenberg (2006) for an introduction and detailed discussions in Aljalian (2005), Mireless (2005).
References


Crommelin, D. J., Storm, G., Verrijkt, R., Leede, L. de, Jiskoot, W., & Hennink, W. E.


Eisenberg, R. S. (2002). Will pharmacogenomics alter the role of patents in drug development? Pharmacogenomics, 3(5), 571-574. (ILL)

477-492.


Gallini, N., & Scotchmer, S. (2002). Innovation policy and the economy. In A. Jaffe,


Guell, R. C. (1997). Haggling for a patent: What a government would have to pay for


Kitch, E. W. (1977, October). The nature and function of the patent system. *Journal of*


Center for Agricultural and Rural Development (CARD) at Iowa State University.


Mossoff, A. (2001). Rethinking the development of patents: An intellectual history 1550-


Markets, Technologies - Institute for Advanced Studies Lucca.


