



Ideas into Practice: How Well Does US Patent Law Implement Modern Innovation Theory?

Stephen M. Maurer, Berkeley Law School and Goldman School of Public Policy, University of California at Berkeley.

Abstract. *The US Supreme Court's decision in Graham v. John Deere (1966) placed neoclassical economic insights at the heart of modern patent law. But economic theory has moved on. Since the 1990s, legal scholars have repeatedly mined the discipline to propose ad hoc rules for individual industries like biotech and software. So far, however, they have almost always ignored the literature's broader lessons for doctrine. This article asks how well today's patent doctrine follows and occasionally departs from modern economic principles.*

The article starts by reviewing what innovation economists have learned since the 1970s. While it is conventional for legal scholars to divide the neoclassical literature into multiple competing "theories," shared mathematical methods and modeling assumptions (e.g. profit-maximization) guarantee that the neoclassical literature's various strands cannot disagree in any fundamental way. For this reason, whatever differences exist in the neoclassical literature are more accurately seen as special cases of a single underlying theory. We argue that this underlying theory must include at least three principles. The first limits reward to non-obvious inventions and explicitly entered the law through Graham's PHOSITA standard. The second principle holds that patent breadth should be chosen to balance the benefits of innovation against the costs of monopoly. Though widely recognized by judges and scholars, the principle's influence on doctrine remains remarkably incoherent. The final principle prescribes rules for allocating patent rewards where multiple inventors contribute to a shared technology. Unlike the first two principles, this insight was unknown in the 1960s and has yet to enter the law.

Remarkably, patent doctrine uses a single concept - Graham's "Person Having Ordinary Skill in the Art" or "PHOSITA" - to address all three principles. This means that doctrinal solutions for one principle can have unintended impacts on the others. In some cases, this link is optional. For example, the article shows how the PHOSITA concept could be generalized to provide separate and distinct tests for, say, non-obviousness and patent breadth. However, other links are mandatory. In particular, the article shows that any doctrinal architecture built on Graham's PHOSITA test automatically allocates reward among successive inventors. Though reasonable, these default outcomes fall short of the economic ideal. The article analyzes how changes in the

Utility, Blocking Patents, Reverse Doctrine of Equivalents, and the Written Description doctrines can mitigate this problem. However, other gaps are inherent and cannot be eliminated without abandoning Graham itself. This radically revised architecture would probably cause more problems than it solves.

Keywords: Patent Law, Patent Incentives, Probabilistic Patents, Patent Infringement, Blocking Patents, PHOSITA, Reverse Doctrine of Equivalents, Written Description Requirement.

JEL Classifications: K19, K41, O34, Z18.

Date: November 28, 2012

Contact: Stephen M. Maurer, smaurer@berkeley.edu.

1. Introduction.¹

Economists' understanding of patent incentives has grown immensely in the four decades since Prof. Kitch wrote his seminal article.² This is clearly a good thing. At the same time, legal scholars have had a hard time mapping these insights into patent doctrine. Since the early 1990s, the usual response has been to divide the economic literature into a half-dozen or so "theories" – the precise number differs from author to author – each of which is said to be valid for some industries but not others.³ But this means asking judges to start every case by selecting whichever theory or combination of theories best fits the facts at hand. This "many theories" approach vests broad discretion in judges and fragments patent doctrine into a potentially endless series of mini-statutes. While economists have argued that an industry-by-industry approach could be efficient,⁴ it is hard to believe that Congress intended such an outcome.

This article begins by arguing that the "many theories" premise is false. Far from being in conflict, most of the economics literature shares common neoclassical⁵ assumptions and methods. For this reason, the putatively separate theories *cannot* disagree in any fundamental way. But in that case, we can hope for a general economic theory leading to universal principles and global legal rules.

¹ The author wishes to thank the Sloan Foundation for its generous support. I also thank Prof. Suzanne Scotchmer for useful comments that very much improved the article. All errors are mine alone.

² Edmund W. Kitch, "The Nature and Function of the Patent System," 20 *J.L. & Econ.* 265 (1977). As with most milestones, assertions that modern innovation economics dates from the 1970s are inherently arbitrary. One might equally argue that the main economic arguments for and against patent law were qualitatively understood in the 1860s. Mark D. Janis, *Patent Abolitionism* (documenting Victorian awareness of the need to protect R&D investment, the danger of "over-stimulated" rewards, and patent thickets); Fritz Machlup and Edith Penrose, "The Patent Controversy in the Nineteenth Century," 10 *J. Econ. Hist.* 1 (1950).

³ For an extended list, see note 11, *infra*.

⁴ Prof. Scotchmer argues that different industries have dramatically different R&D costs and/or ability to extract revenue from consumers. These factors make it "almost inevitable" that a one-size-fits-all patent reward will under-reward some inventors and over-reward others. Suzanne Scotchmer, *Innovation and Incentives* (MIT Press 2004) at 117-118.

⁵ The term "neoclassical economics" conventionally refers to "approaches to economics focusing on the determination of prices, outputs, and income distributions in markets through supply and demand, often mediated through a hypothesized maximization of utility by income-constrained individuals and of profits by cost-constrained firms employing available information and factors of production, in accordance with rational choice theory." Wikipedia, "Neoclassical Economics," http://en.wikipedia.org/wiki/Neoclassical_economics#cite_note-0; see also, Antonietta Campus, "Marginal Economics," *The New Palgrave: A Dictionary of Economics*, v. 3, p. 323 (1987).

This article asks how well current patent law implements the lessons of neoclassical theory. In particular, we argue that any sufficiently unified neoclassical theory of innovation incentives must endorse three specific policy goals: (1) Limiting reward for inventions that would be discovered in any case; (2) setting rewards that balance the benefits of innovation against the burden of monopoly prices and wasteful duplication of effort; and (3) allocating reward to cover each innovator's costs when successive advances contribute to a common technology. Remarkably, we will see that current patent doctrine uses a single concept – *Graham's* "Person Having Ordinary Skill in the Art" – to address all three goals. Inevitably, this creates problems. We show that any version of patent law that uses a *Graham* non-obviousness threshold to address our first goal *automatically* sets "default choice" rules for reaching our third goal. Though reasonable, these rules fail to implement the ideal incentives recommended by neoclassical theory. In what follows, we show that the fit can be improved by adding defenses (*e.g.* the Reverse Doctrine of Equivalents) that require jurors to perform additional fact inquiries. That said, no amount of tinkering can eliminate the discrepancy entirely. This implies that *any* doctrine built on *Graham* non-obviousness is inherently flawed and that truly fundamental reforms would – assuming that they are possible at all – require fundamental changes to current law.

We proceed as follows. Section 2 discusses how innovation economics research has entered the legal literature. It argues that the "many theories" viewpoint has been overstated and that most of the literature falls within a coherent "neoclassical core." Section 3 uses this neoclassical core to identify three policy goals that any economically efficient patent statute must try to accomplish. Section 4 reviews the Utility doctrine and compares it against modern economic insights. Section 5 explores a puzzle: R&D costs are ubiquitous in economic theory but seldom if ever mentioned in doctrine. We argue that patent law overcomes this difficulty by using the PHOSITA concept to measure inventive cleverness as a proxy for effort and cost. We also discuss how litigation uncertainty improves the match between doctrine and economically efficient incentives. Section 6 analyzes the allocation rules generated by *Graham's* PHOSITA test in a simple Literal Infringement scenario. Strikingly, we find that the test generates rewards that are facially reasonable but highly imperfect. Section 7 extends the analysis to the Doctrine of Equivalents. Section 8 examines how defenses based on the Reverse Doctrine of Equivalents, Pioneer Patents, and Written Description defenses can improve the alignment between doctrine and economically efficient incentives. Section 9 assesses the gaps between patent doctrine and our three policy goals and asks whether a fundamentally different architecture could do better. Section 10 presents a brief conclusion.

2. The Search for a General Theory

No one doubts that a general theory of patent law is desirable. How could they? We live in an age where general theories – sometimes even “Theories of Everything”⁶ – are praised to the skies. The only real question is whether we know how to build such a thing.

This Section begins by asking what we mean by a general theory. For the sake of definiteness, we illustrate our argument with paradigmatic examples from physics. We then review various “theories” that legal scholars have identified in the economics literature. Significantly, most of this work shares the same neoclassical assumptions (rational actors, profit maximization) and mathematical methods (optimization calculus). Within this “neoclassical core,” any disagreements are superficial and can usually be traced to special industry-specific facts. From this perspective, the putatively separate “theories” are more accurately seen as special cases of a common underlying theory. This should encourage patent law reformers to look for broad principles and rules of general application.

A. The Gold Standard: Classical Physics

We begin by recalling what a really powerful theory looks like. Here the paradigmatic example is physics. Consider the path of a moving object. Plainly, one can imagine many different versions of this problem. For example we can analyze the object’s path when it is allowed to move freely in three dimensions like a planet orbiting the sun (Problem A). Alternatively, we can analyze the much simpler problem of how fast it moves when forced to follow a track or inclined plane. (Problem B). Now both problems were solved a long time ago, and the two solutions look different on the page.⁷ The question is, how should we talk about them? Or, better still, how is it *useful* to talk about them? One could say that Problem A is a “different theory” than Problem B. But that distinction is misleading. Instead, we usually say that B is “special case” of the “general theory” A.

The difference is more than semantic. Problem B, being less general, is almost always easier to solve than Problem A. Indeed, this is usually why Problem B was chosen in the first place. On the other hand, each special case identifies features that must also appear in the general theory. For a physicist, finding this general solution is usually an end in itself. However, general solutions also offer power over the world. For example, engineers might use a really general

⁶ See, e.g., Steven W. Hawking, *The Theory of Everything* (Jaiko 2007)

⁷ Compare Richard P. Feynman, Robert B. Leighton and Matthew Sands, *The Feynman Lectures on Physics* (Addison Wesley 1977) at pp. 8-8 to 8-10 (ball rolling down inclined plane) with pp. 9-6 to 9-9 (planetary orbits).

solution to plot lowest-energy trajectories for spacecraft traveling between any two points in the solar system.

Clearly, finding these solutions is hard. One problem is that physicists may not know the general theory.⁸ Strangely, this is seldom fatal. Instead, it has often been enough to suspect that a general theory exists and guess at principles.⁹ On the other hand, some general theories are known but lack general solutions.¹⁰ Even if a general theory exists, then, engineers may still be restricted to special (and solvable) cases.

The corresponding agenda for patent law seems clear. Economists, like physicists, are groping toward a general theory of innovation. And lawyers, like engineers, would like to exploit this knowledge to design doctrines that accelerate innovation at the lowest possible cost. As with our physics example, we cannot know in advance whether this agenda is possible. At the same time, we can be sure of one thing – We will never know unless we try. This article begins the experiment.

B. How Lawyers See the Economics Literature

It has become conventional for legal scholars to stylize the economics literature in terms of competing theories.¹¹ The practice seems to have originated in a 1990 article by Profs. Merges

⁸ This was the case for electromagnetism between 1831 and 1865, and relativity between 1905 and 1917. *See, e.g.*, Roy Porter, *The Biographical Dictionary of Scientists* (Oxford Univ. Press 1994) (2d ed.) (entries for Faraday, Maxwell, and Einstein).

⁹ For example, Maxwell's final theory of electromagnetism includes one law previously known to Faraday and a second law that Ampere discovered and Maxwell corrected. Wikipedia, "Maxwell's Equations," [http://en.wikipedia.org/wiki/Maxwell's Laws#Ampere's Law with Maxwell's correction](http://en.wikipedia.org/wiki/Maxwell's_Laws#Ampere's Law with Maxwell's correction).

¹⁰ The legend of Mohammed's Coffin provides an entertaining example. According to medieval folklore, the Prophet Mohammed was buried in such a way that magnets kept his coffin perpetually suspended between heaven and earth. Samuel Earnshaw proved in 1842 that no general solution to the Mohammed's coffin problem exists. There are, however, solutions for special cases including the one-dimensional problem and time-varying electric fields. Wikipedia, "Earnshaw's Theorem," [http://en.wikipedia.org/wiki/Earnshaw's theorem](http://en.wikipedia.org/wiki/Earnshaw's_theorem). A lawyer would call these solutions "*ad hoc*" rules.

¹¹ *See, e.g.*, Lisa Larrimore Ouellette, "Do Patents Disclose Useful Information?" 25 *Harv. J. Law & Tec.* 531 (discussing disclosure, reward, prospect, and racing theories) (2012); Amelia S. Rinehart, "Patents as Escalators," 14 *Vand. J. Ent. & Tech. L.* 81 (2011) (discussing reward, prospect, and option pricing theories); Robert P. Greenspoon and Catherine M. Cottle, "Don't Assume a Can Opener: Confronting Patent Economic Theories With Licensing and Enforcement Reality," 12 *Colum. Sci. & Tech. L. Rev.* 194 (2011) (discussing reward, prospect, and commercialization theories); Amir H. Khoury Differential Patent Terms and the Commercial Capacity of Innovation 18 *Tex. Intell. Prop. L.J.* 373, 400 at n. 139 (2010) (discussing reward, disclosure, commercialization, and broad prospect theories); Ted Sichelman Commercializing Patents 62 *Stan. L. Rev.* 341 (2010) (discussing prospect, reward, commercialization, signaling, and transaction cost theories); Andrew Blair Stanek, "Increased Market Power as a New Secondary Consideration in Patent Law," 58 *Am. U.L. Rev.* 707 (2009) at n. 136 (identifying five

and Nelson, who criticized earlier scholars for “assum[ing] that invention is the same in all technologies.”¹² They divided the economics literature into three categories. First, there were traditional models that examined incentives for a single inventor working in isolation. These almost always focused on “a simple tradeoff” between “incentives to the inventor and underuse of the invention due to patent monopolies.”¹³ Second, there were more recent

economic theories posited over time for the patent system); Wolrad Prinz zu Waldeck und Pyromont, “Research Tool Patents After *Integra v. Merck* – Have They Reached a Safe Harbor?” 14 *Mich. Telecomm. Tech. L. Rev.* 367 (2008) (discussing inducement-to-invent, incentive to invest, reward, disclosure, and anticommons theories); Carol M. Nielsen and Michael R. Samdzija, “Compulsory Patent Licensing: Is It a Viable Solution in the United States?” 13 *Mich. Telecomm. Tech. L. Rev.* 509, 517-520 (2007) (discussing prospect, cumulative innovation, patent thicket, and anticommons theory); David B. Conrad, “Mining the Patent Thicket: The Supreme Court’s Rejection of the Automatic Injunction Rule in *eBay v. MercExchange*,” 26 *Rev. Litig.* 119, 132 – 137 (2007) (discussing reward, disclosure, commercialization, and prospect theories); F. Scott Kieff, “Coordination, Property and Intellectual Property: An Unconventional Approach to Anticompetitive Effects and Downstream Access,” 56 *Emory L.J.* 327 (2006) (discussing “reward,” “rent dissipation,” “anticommons,” “prospect,” and “commercialization” theories); Richard S. Gruner, “Corporate Patents: Optimizing Organizational Responses to Innovation Opportunities and Invention Discoveries,” 10 *Marq. Intell. Prop. L. Rev.* 1 (2006) (discussing reward, prospect, disclosure, and rent dissipation theories); Peter Fox, “It’s Not Over the for the Product of Nature Doctrine Until the Synthetic Super-Heavy Element (‘SHE’) Sings,” 79 *Temp. L. Rev.* 1005 (2006) at 1011 and n. 47 (discussing reward, prospect, race-to-invent, and rent dissipation theories); Timothy R. Holbrook, “Possession in Patent Law,” 59 *SMU L. Rev.* 123 (2006) (discussing reward, portfolio, prospect, incentive, and signaling theories); Christopher A. Cotropia, “Patent Claim Interpretation Methodologies and Their Claim Scope Paradigms,” 47 *Wm. & Mary L. Rev.* 49, 117-124 (2005) (discussing competitive innovation, prospect theory, and cumulative innovation); Shubha Ghosh, “Patents and the Regulatory State: Rethinking the Patent Bargain Metaphor After *Eldred*,” 19 *Berkeley Tech. L.J.* 1315 (2004) (discussing prospecting, information costs, business asset, and industry regulation theories of patent law); Dan L. Burk & Mark A. Lemley, “Policy Levers in Patent Law,” 89 *Va. L. Rev.* 1575, 1600-1614 (2003) (discussing prospect, competitive innovation, cumulative innovation, anticommons, and patent thicket theories); Julian David Forman, “A Timing Perspective on the Utility Requirement in Biotechnology Patent Applications,” 12 *Alb. L.J. Sci. & Tech.* 647 (2002) (discussing “prospect,” “reward,” and Anticommons theory); Mark. D. Janis, “Second Tier Patent Protection,” 40 *Harv. Int’l L.J.* 151 (1999) (discussing prospect, reward, and anticommons theories); Phanesh Koneru, “To Promote the Progress of Useful Articles? An Analysis of the Current Utility Standards of Pharmaceutical Products and Biotechnology Research Tools,” 38 *IDEA* 625 (1998) (discussing reward, patent-induced, prospect, race-to-invent, and rent dissipation theories); Craig Allen Nard, “Deference, Defiance, and the Useful Arts,” 56 *Ohio St. L.J.* 1415, 1419 & n. 13 (1995) (discussing reward, disclosure, prospect, and rent dissipation theories).

Probably the most systematic “many theories” analysis is found in A. Samuel Oddi, *Un-Unified Economic Theories of Patents-the-Not-Quite-Holy Grail*, 71 *Notre Dame L. Rev.* 267 (1996). Prof. Oddi identifies five economic theories that have been developed to analyze patent incentives and asks if any of them supply a “unifying theory of patent law” *Id.* 268. Here, the phrase “of patent law” is crucial. Unlike the present article, Oddi is emphatically *not* looking for a unified theory of innovation that might or might lead to ideal patent rules. Rather, he asks whether *existing* doctrine can be “explained” by any single theory. *Id.* at 321. This means, among other things, rejecting theories that are not “consistent with the [current] patent statute, its substance, and procedure.” *Id.* at 271. Not surprisingly, the quest fails: Indeed, Oddi concludes that some decisions are “better explained by doctrine than by economic theory.” *Id.* at 290.

¹² Robert P. Merges and Richard R. Nelson, “On the Complex Economics of Patent Scope,” 90 *Colum. L. Rev.* 839, 843 (1990).

¹³ *Id.* at 869.

theories in which innovation proceeded on the basis of “many actors” operating over “many rounds.”¹⁴ Following Kitch, these often produced arguments in favor of broad patents that deliberately promote “single-firm domination of a technological prospect.”¹⁵ Third, and finally, Merges and Nelson argued that the usual neoclassical assumption that firms are rational and maximize profits could fail. In this case, Kitch-type arguments disappear so that narrow patents are once again preferred.¹⁶

Scholars have updated Merges and Nelson’s categories over the years. In their influential 2003 article, Profs. Burk and Lemley subdivide Merges and Nelson’s second (multi-inventor) category into “Prospect Theory,” “Competitive” innovation, and “Cumulative Innovation.”¹⁷ Similarly, they divide the third category into “Anticommons Theory,” and “Patent Thickets.”¹⁸ Despite small variations from author to author, this expanded list continues to dominate contemporary academic discussion.¹⁹

Inevitably, this “many theories” viewpoint implies a choice. In the easy cases, one published theory will indeed fit all of the available facts. This, however, can only happen by accident. As Merges and Nelson acknowledge, individual industries more typically display “some attributes” of the “true” published theory while “lack[ing] all the features.”²⁰ They may also have “particular characteristics of [their] own” that do not fit any theory at all.²¹ Logically, judges can address such cases in just two ways: Pick an existing category knowing that the fit is imperfect or else let the number of recognized “theories” proliferate. While Merges and Nelson do not address the issue in detail, they are plainly open to the latter choice. Despite arguing that four

¹⁴ *Id.* at 869.

¹⁵ *Id.* at 871-72.

¹⁶ Merges and Nelson argue, in particular, that real firms may not be satisfied with sub-optimal performance, have organizational biases that limit their rationality, and/or be unable to agree on probable outcomes even in principle. *Id.* at 872-873.

¹⁷ Dan L. Burk & Mark A. Lemley, “Policy Levers in Patent Law,” 89 *Va. L. Rev.* 1575 at pp. 1600-1610 (2003).

¹⁸ *Id.* at 1610-1615.

¹⁹ See note 10, *supra*.

²⁰ Merges & Nelson, *supra*, at 882-883; *cf.* Dan L. Burk & Mark A. Lemley, “Biotechnology’s Uncertainty Principle,” 54 *Case W. Res. L. Rev.* 691, 738 (2004) (arguing that biotech industry is “properly described in part by the anticommons theory ... and in part by prospect theory ...”),

²¹ *Id.* at 883.

categories cover much of the US economy,²² they admit that still more categories could be needed in the future.²³ Worse, they remark that the correct theory can differ from firm-to-firm and even product-to-product.²⁴

This modern trend marks a fundamental departure from patent law's historic emphasis on rules of general application.²⁵ Despite this, modern scholars continue to treat policy questions as an exercise in picking and choosing between economic theories.²⁶ Judges seem to be similarly fond of inventing industry-specific rules.²⁷

C. Searching for Unity: The Neoclassical Core

Once legal scholars decided on a “many theories” approach, the rise of industry- and sometimes even product-specific rules became inevitable. But was the decision necessary? In what follows, we review the economic literature's main strands and find evidence for their underlying unity.

²² The four categories are “discrete innovations” (chemical compounds), “cumulative technologies” (electric lighting, automobiles, aircraft, radio, semiconductors and computers), “chemicals” (bulk chemicals and pharmaceuticals), and “science-based industries” (biotech, medical diagnostics, and lasers). Merges and Nelson, *supra* at 884-908.

²³ Merges and Nelson only argue that the US economy is adequately described by “at least” four categories. *Id.* at 880.

²⁴ For example, Merges and Nelson argued that chemical products with a single application are “discrete products,” but chemical products with multiple applications are not. Additionally, processes for *making* chemicals are “cumulative” technologies. *Id.* at 882-83.

²⁵ Samantha A. Jameson, “The Problems of Utility Analysis in *Fisher* and Its Associated Policy Implications and Flaws,” 56 *Duke LJ* 311, 339-41 (2006) (noting that patent system “theoretically” provides “technology-neutral protection to all kinds of technologies”); Dan L. Burk & Mark A. Lemley, “Is Patent Law Technology-Specific?” 17 *Berkeley Tech. L.J.* 1155, 1160 (2002) (noting that patent law traditionally stressed “unified rules” for a “homogeneous world”).

²⁶ See, e.g. Alan Devlin, “Patent Law's Parsimony Principle,” 25 *Berkeley Tech. L.J.* 1693, 1721, 1738-39 (2011) (arguing that patent rules should be adjusted to correct “over-rewarding” of inventors in the business methods, computer software, and medical diagnostics industries); Michael A. Carrier, “Cabining Intellectual Property Through a Property Paradigm,” 54 *Duke L.J.* 1, 38-42 (2004) (identifying eleven industries where “patents are less critical for innovation than the traditional theory would posit” implying the need for narrower patents and/or a “greater role for defenses to claims of patent infringement”).

²⁷ For example, Profs. Burk and Lemley claim that the Federal Circuit goes to “inordinate lengths” to find biotech inventions non-obvious while simultaneously imposing stringent enablement and written description requirements. The situation in software is exactly the opposite. Dan L. Burk & Mark A. Lemley, “Policy Levers in Patent Law,” 89 *Va. L. Rev.* 1575, 1593 (2003).

We begin by examining the literature on individual inventors. This genre is variously called **reward**,²⁸ **incentive to invent**,²⁹ and **inventive**³⁰ theory. Regardless, the central focus is the same: Adjusting the size of the patent reward for maximum benefit to society. On the one hand, the reward must usually be high enough to cover inventors' R&D costs³¹ and may require a bonus to accelerate R&D or overcome the probability of failure or unexpected obstacles. On the other, new products are pointless unless consumers can afford them. Making the patent monopoly too broad interferes with this goal.³² While these points were already well-known to Thomas Jefferson,³³ modern neoclassical theory makes them precise. Most legal scholars have absorbed this more careful and sophisticated language.³⁴

So far, we have concentrated on theories that address inventors in isolation. However, real inventors often coexist and compete with one another. Prof. Kitch's **prospect theory**³⁵ was the

²⁸ For recent examples, see, e.g., Lisa Larrimore Ouellette, "Do Patents Disclose Useful Information," 25 *Harv. J. Law & Tec* 531, 541 (2012); Amelia S. Rinehart, "Patents as Escalators," 14 *Vand. J. Ent. & Tech. L.* 81, 89 and note 42 (2011); Robert P. Greenspoon and Catherine M. Cottle, "Don't Assume a Can Opener: Confronting Patent Economic Theories With Licensing and Enforcement Reality," 12 *Colum. Sci. & Tech. L. Rev.* 194, 196 (2011).

²⁹ Teresa M. Summers, "The Scope of Utility in the Twenty-First Century: New Guidance for Gene-Related Patents," 91 *Geo L.J.* 475, 485-488 (2003). The usage has the advantage of emphasizing economic incentives over natural law arguments based on justice to the inventor. A. Samuel Oddi, "Un-unified Economic Theories of Patents--The Not-Quite-Holy Grail," 71 *Notre Dame L. Rev.* 267, 275-78 (1996) (distinguishing reward theories based on economic incentives from those based on natural law); Yusing Ko, "An Economic Analysis of Biotechnology Patent Protection," 102 *Yale L.J.* 777, 791-793 (1992).

³⁰ Laurie L. Hill, "The Race to Patent the Genome: Free Riders, Hold Ups, and the Future of Medical Breakthroughs," 11 *Tex. Intell. Prop. L.J.* 221, 238 (2003). Scholars sometimes note a fourth **incentive to disclose** theory which holds that patents are needed to persuade inventors to publish information that would otherwise be kept secret. Summers, *supra* 488-491; Hill, *supra* at 238; Yusing Ko, "An Economic Analysis of Biotechnology Patent Protection," 102 *Yale L.J.* 777, 795-798 (1992); Liza Vertinsky, "An Organizational Approach to the Design of Patent Law," 13 *Minn. J.L. Sci & Tech.* 211, 221 (2012).

³¹ Unless otherwise stated, we define the "cost" of an invention as the minimum amount of money needed to conduct a particular R&D project.

³² See, e.g. Kitch, at 266; Ko, *supra*, note 28 at 792.

³³ See *Graham v. John Deere*, 383 U.S. 1 (1966). See also, Summers, *supra*, at 485-488 (2003) (noting that incentive to invent, incentive to disclose, and incentive to innovate theories "have been advanced for centuries to justify the patent system").

³⁴ Arti Kaur Rai, "Regulating Scientific Research: Intellectual Property Rights and the Norms of Science." 94 *N.W. U. L. Rev.* 77, 116-117 (1999) (arguing that patents internalize the positive externalities of innovation and discourage free-riding); Forman, *supra* at 663 (identifying "traditional view of the patent bargain" in terms of free ridership, public goods, and monopoly pricing); Hill, *supra* at 238 (freeriding may delay invention).

³⁵ Kitch, *supra*. In contrast to later commentators, Kitch himself did not adopt a "many theories" viewpoint. *Id.* at 266 ("The reward theory is not questioned on its own terms. Rather it is argued that the reward theory offers an incomplete view of the patent system.")

earliest installment in this multi-inventor literature. It argued that multiple, uncoordinated inventors were bound to duplicate each others' work and that this would necessarily cause duplication and waste. Broad patents overcome this problem by letting a single owner coordinate development.³⁶ Variants of Kitch's argument add that patent owners can also use their power to suppress rent-seeking³⁷ and promote information-sharing.³⁸ Many scholars argue that these ideas are most relevant for R&D programs that "commercialize" primitive prototypes into marketable products. For this reason, Kitch's basic idea often appears in the guise of "development,"³⁹ "incentive to innovate, collaborate, and commercialize,"⁴⁰ and "innovation theory."⁴¹

Kitch's theory was based on accidental duplication, *i.e.* the idea that non-interacting inventors would inevitably duplicate each others' work. **Racing** theory extends this idea by pointing out that the case for interacting inventors is even worse. Patent law, after all, only rewards the *first* inventor while second-comers get nothing. The resulting patent races can be beneficial to the extent that they accelerate innovation and provide redundancy against failure. But the drive to be first – even by an hour or two⁴² – can also force competing R&D programs deep into diminishing returns. This, of course, is wasteful. As in the classical reward literature, innovation theory teaches that policymakers should pick whichever reward level offers the greatest net benefit.

³⁶ Lemley & Burke, *supra*, 89 *Va. L. Rev.* at 1600; Koneru, *supra*, 38 *IDEA* at 634.

³⁷ Forman, *supra*, at 665; Kenneth W. Dam, "The Economic Underpinnings of Patent Law," 23 *J. Legal Stud.* 247, 253 (1994).

³⁸ Ko, *supra* at 800-802 (1992).

³⁹ Rai, *supra*, at 120-121; *see also* Forman, *supra* at 665 (arguing that Kitch's theory is focused on "the further development of prospects").

⁴⁰ Summers, *supra* at 491-493; *see also*, Ko, *supra* at 799 (incentive to innovate).

⁴¹ Hill, *supra* at 238. Hill somewhat confusingly divides prospect theory (suppressing wasteful and duplicative investments) from innovation theory (covering commercialization costs). *See also* Ko, *supra*, at 791-804. Vertinsky argues that Kitch's prospect theory is more usefully seen as an aspect of commercialization. Liza Vertinsky, "An Organizational Approach to the Design of Patent Law," 13 *Minn. J.L. Sci & Tech.* 211 (2012).

⁴² American folklore holds that Alexander Graham Bell filed his patent application for the telephone "an hour or two" before his nearest rival. Wikipedia, "Elisha Gray and Alexander Bell Telephone Controversy," http://en.wikipedia.org/wiki/Elisha_Gray_and_Alexander_Bell_telephone_controversy#First_to_arrive_at_the_patent_office. Since US patent law used a first-to-invent rule, this probably did not matter from a legal standpoint. Even so, the story reminds us how much can turn on being a minute or two faster than competitors.

Finally, **cumulative innovation** addresses how the patent reward should be divided when multiple inventors contribute to a common technology.⁴³ Here, the key insight is that a first generation (“1G”) invention offers two distinct benefits to society. First and most obviously, it yields *present value* by telling society how to make products for immediate use. Second, it delivers *option value* by providing a platform that second generation (“2G”) inventors can build on to invent still more products. Ideally, inventors should proceed whenever *all* benefits *including* option value exceed total costs. This cannot happen, however, so long as each inventor’s reward is limited to the revenues earned by his or her products. In this case, inventors who produce intermediate (pre-commercial) inventions receive nothing for option value so that the technology is never developed at all. Correcting this problem means allocating part of the 2G (3G, 4G, 5G...) inventor’s revenues back to her 1G (2G, 3G, 4G,...) forebears. We return to this problem below.

So far we have stressed the differences between our different neoclassical innovation models. These invariably involve contingent facts (one inventor vs. many, non-interacting inventors vs. racing) that differ from industry to industry. What really matters, though, are the commonalities. First, each of the foregoing literatures assumes that firms maximize profits. Second, each assumes that actors pursue this goal rationally. Third, each invokes the same methodology – supply, demand, and occasionally game theory. Finally, all depend on the same mathematical machinery of calculus and optimization. These commonalities guarantee that the putatively separate theories cannot disagree in any deep sense. In keeping with our physics discussion, we should instead think of them as fact-dependent special cases. This encourages us to look for unifying principles and rules of general application.⁴⁴

D. Beyond Neoclassical Theory

In keeping with the legal literature’s conventional categories, we end with theories that reject the assumptions, methods, and even goals of neoclassical economics. These have enjoyed something of a renaissance in the two decades since Merges and Nelson wrote their article.

⁴³ See, e.g. Suzanne Scotchmer, *Innovation and Incentives* (MIT Press 2004) at pp. 127-159; Jerry R. Green and Suzanne Scotchmer, "On the Division of Profit in Sequential Innovation," *RAND Journal of Economics*, 26(1): 20-33 (1995); S. Scotchmer, "Standing on the Shoulders of Giants: Cumulative Research and the Patent Law," *Journal of Economic Perspectives* 5:29-41 (1991); Ofer Tur-Sinai, "Cumulative Innovation in Patent Law: Making Sense of Incentives," 50 *IDEA* 723 (2010).

⁴⁴ This situation is normal in law, where legislators and judges almost always act on less-than-perfect information.

The Anticommons concept is far the biggest development.⁴⁵ Though often denoted a theory,⁴⁶ a close reading of Prof. Heller and Eisenberg's seminal paper shows that the idea is really a conjecture that – naïve neoclassical predictions notwithstanding – many worthwhile transactions never take place. Heller and Eisenberg then offer three reasons (“structural concerns”) why this could happen. The first (“transaction costs” and “strategic behaviors”) notes that even neoclassical theory predicts occasional deadlock.⁴⁷ Examples include so-called mixed strategy games in which players veto profitable agreements in hopes of being bought off⁴⁸ and the familiar dynamic that forces owners of monopoly complements to set higher prices than a cartel would.⁴⁹

If this were all, one could still hope to unify the Anticommons with neoclassical theory. However, Heller and Eisenberg's next two concerns make this impossible. The first denies profit maximization. All would be well, the authors argue, if every patent owner was trying to maximize profit. But suppose the parties pursue different and perhaps incompatible goals like a desire to hurt rivals?⁵⁰ In such a world, conventional microeconomic assumptions no longer hold. Heller and Eisenberg's final critique is even more fundamental. Neoclassical theory assumes rationality. However, “cognitive biases” could lead actors to “systematically overvalue their assets and disparage the claims of their opponents.” Taken literally, this implies that no two parties can ever agree on an invention's value.⁵¹

⁴⁵ Michael A. Heller, “The Tragedy of the Anticommons: Property in the Transition from Marx to Markets,” 111 *Harv. L. Rev.* 621 (1998); Michael A. Heller, “The Boundaries of Private Property,” 108 *Yale L. J.* 1163 (1999); Michael A. Heller and Rebecca S. Eisenberg, “Can Patents Deter Innovation? The Anticommons in Biomedical Research,” *Science* 280(50):698-701 (1998) (extending Anticommons concept to patent law). For a more recent and extended statement, see Michael Heller, *The Gridlock Economy: How Too Much Ownership Wrecks Markets, Stops Innovation, and Costs Lives* (Basic Books: 2010).

⁴⁶ See, e.g. Robert P. Merges, “Autonomy and Independence: The Normative Face of Transaction Costs,” 53 *Ariz. L. Rev.* 145 at 145 (2011).

⁴⁷ Michael A. Heller and Rebecca S. Eisenberg, *supra* at 698.

⁴⁸ For example, Prof. Lemley notes that rightsholders may play a game of chicken in which each argues that he is “irrational enough to kill the whole deal unless you get more than your ‘fair’ share.” He then adds that this can “result in the parties failing to come to terms at all.” Mark A. Lemley, “The Economics of Improvement in Intellectual Property Law,” 73 *Tx L. Rev.* 989, 993 (1997).

⁴⁹ See, e.g. Scotchmer, *supra*, at pp. 142-146.

⁵⁰ Michael A. Heller and Rebecca S. Eisenberg, *supra* at 700-701. For example, Prof. Lemley remarks that a government agency might try to maximize public access, an author might try to suppress hostile reviews, and a business might want to hurt rivals. Lemley, *supra*, 73 *Tx L. Rev.* at 1061.

⁵¹ Heller and Eisenberg, *supra* at 701.

Given that so much of the Anticommons concept is based on rejecting neoclassical assumptions, it is hard to see how any kind of unification is possible. Still, one could imagine a situation where judges only ignored neoclassical predictions in cases where Anticommons effects are strong. But when, exactly, is that? For transactions involving two negotiators? For ten? For one hundred? The Anticommons literature does not say.

Patent Thickets. Neoclassical theory works best when transaction costs are small and can be safely neglected. This condition is violated when patent searches are prohibitively expensive.⁵² The question is, what then? Within the neoclassical framework, economists usually invoke “transactions cost” arguments as a reason to organize existing institutions differently. But suppose that the patent system is already (nearly) as efficient as it can be. Then transactions costs are unavoidable and we can imagine three distinct situations:

Small Transactions Costs. If patent thicket effects are unimportant, policymakers can ignore them.

Moderate Transactions Cost. If transactions costs are truly unavoidable, we must include them in the product’s R&D cost. The fact that the costs involve, say, lawyers instead of bench scientists is immaterial. On the other hand, society should never invest more in an R&D project than it hopes to gain from innovation. From this standpoint, we should be *grateful* if transactions costs stop R&D projects from going forward.

Large Transactions Costs. It is possible to imagine a world where transaction costs make all R&D projects prohibitively expensive. Here, patents yield few benefits and should be abolished.

The problem, for now, is that the Patent Thicket literature – like the Anticommons – says almost nothing about how to identify these cases in practice.

Other Theories. One can, of course, gin up additional theories indefinitely. In practice, this usually means suspending particular neoclassical assumptions. Prof. Vertinsky has compiled an extended list of how non-neoclassical theories based on bounded rationality, imperfect information, and/or opportunism could potentially improve doctrine.⁵³ Strikingly, however,

⁵² T Randolph Beard and David L. Kaserman, “Patent Thickets, Cross-Licensing, and Antitrust” *Antitrust Bulletin* 345-368 (Fall 2002).

⁵³ Liza Vertinsky, “An Organizational Approach to the Design of Patent Law,” 13 *Minn. J.L. Sci & Tech.* 211, 221 (2012); see also, Molly Shaffer Van Houweling, “Author Autonomy and Atomism in Copyright Law,” 96 *Va. L. Rev.* 549, (2010) (asserting that proliferation of works subject to copyright, number of people owning copyrights, and type/size of separately owned rights have complex and often contradictory impacts on author autonomy, publishers’ ability to license, competition, diversity of expression, and distributive fairness.)

almost all of these insights are narrow or ambiguous. While they may sometimes extend neoclassical theory, they cannot replace it.

Natural Law. So far, we have focused on theories that reject neoclassical assumptions. However, the goal of neoclassical theory – maximizing society’s total wealth – is also a choice. If we want to, we can set different goals. For natural law, this usually means justice to the inventor.⁵⁴

The problem, once again, is predictive power. Natural law provides a reasonably precise theory of why someone who creates a physical object should own it. But re-potting this idea in patent law leads to difficulties. How broad should patent right be? Should the patent right (unlike physical property) expire? And if it does, how long should temporary ownership last? Natural law does not say.

E. Privileging The Neoclassical Core

Neoclassical theory is remarkably unified and coherent. We argue in the next section that these qualities provide clear and determinate guidance. This provides obvious practical advantages for legal scholars. Still, this clarity comes at a cost. Many non-neoclassical theories – some of which routinely appear in the legal literature – will have to be discarded. We should only do this if unifying these theories with neoclassical innovation economics would destroy the latter’s clarity and usefulness.

The obstacles to unification are deepest for natural law. The basic difficulty is that the elegance and power of neoclassical theory depends on a very special (“Pareto optimal”) definition of economic efficiency. But natural law starts from a very different (“fairness to the inventor”) goal. Assuming that it could be constructed at all, any theory that combines utility and natural law is bound to be clumsy and indeterminate. And of course, American law has already made its choice. As Prof. Lemley remarks, the “primacy of incentive theory in justifying intellectual

⁵⁴ For a modern version of these arguments, see Robert P. Merges, *Justifying Intellectual Property* (Harvard Univ. Press 2011). A detailed account of the 19th Century debate can be found in Janis, *supra*.

property” is anchored in the language of the Constitution’s Patent and Copyright Clause⁵⁵ as interpreted by case law.⁵⁶

The prospects for unification within other, non-neoclassical branches of economics are brighter. Indeed, economists routinely construct non-neoclassical theories.⁵⁷ The problem, for now, is that the Anticommons and patent thickets are not yet theories in this sense. Should judges start to worry when the number of overlapping patents is two, or eight, or fifty? Are there simple institutional arrangements – for example cross-licenses – that can overcome the Anticommons? The literature does not say.

In many ways, the situation is similar to antitrust law, which invokes neoclassical economics to offer a single cure (“more competition”) to the world’s ills. Strikingly, judges do this even though microeconomic theory *acknowledges* that competition can sometimes lead to disaster. As the Supreme Court emphasized in *Trinko*, such evils as do occur are best left to other bodies of law.⁵⁸ This article argues that patent doctrine should similarly build on the underlying coherence of neoclassical theory. To the extent that neoclassical assumptions fail, policymakers can always fall back on tax policy, sponsored research, and other non-patent incentives.

3. Three Policy Goals

The fact that neoclassical theories are logically consistent encourages us to think that we can write rules of general application. This Section begins the process by extracting three policy goals that appear throughout the neoclassical literature.

⁵⁵ US Constitution Art. I, §8. (“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.”)

⁵⁶ Mark A. Lemley, “The Economics of Improvement in Intellectual Property Law,” 73 *Tx L. Rev.* 989, 993 (1997); see also Merges and Nelson, *supra*, at 852 (“While there are many who may challenge the propriety of these goals from the standpoint of economic policy, it is both realistic and necessary for courts and the patent office to pursue the goals implicit in both constitutional and statutory provisions.”); Liza Vertinsky, “An Organizational Approach to the Design of Patent Law,” 13 *Minn. J.L. Sci & Tech.* 211, 221 (2012) (“[T]he traditional justification for the patent regime in the United States has been largely a utilitarian one based on the public goods aspect of innovation”).

⁵⁷ See, Vertinsky, *supra*.

⁵⁸ *Verizon Comms. Inc. v. Law Offices of Curtis V. Trinko*, 540 US 398 (2004).

A. Goal No. 1: Avoiding Redundant Incentives.

Patents are superfluous for products that would be invented anyway.⁵⁹ This could happen because inventors can gain temporary monopolies (“first mover advantages”) even without formal patents or because they already possess market power over related goods. Additionally, society offers many incentives besides patents. These include prizes, grants, and subsidies. Finally, the rise of open source communities reminds us that human beings are inherently creative.⁶⁰ This basal rate of innovation would presumably continue even if patents disappeared entirely.⁶¹ In all of these cases, granting a patent imposes a tax on innovation to no purpose.

This suggests that the patent system does not have to cover every inventor’s full R&D cost. The problem is deciding where to draw the line. Profs. Cohen, Nelson, and Walsh have collected survey evidence showing that CEOs in most industries see patent incentives as relatively unimportant. However, their work also shows that the answer varies by industry.⁶² Evidently, facts matter. By itself, theory is helpless to say which inventions should and should not be patentable.

The Supreme Court famously addressed this problem in the 1960s. Its *Graham*⁶³ decision acknowledged the importance of “drawing a line between the things which are worth to the public the embarrassment of an exclusive patent, and those which are not.”⁶⁴ It also announced a test: Inventions should be patentable if and only if they were “non-obvious” to a “Person Having Average Skill in the Art” (*i.e.*, a “PHOSITA”). We argue below that this was a very useful concept. At the same time, the Court failed to articulate explain why it put the

⁵⁹ Koneru, *supra*, p. 632 (patents should not be granted where inventions would would be developed anyway “for reasons unrelated to the existence of the patent grant.”); Summers, *supra*, at 486 (noting that “sufficient incentives for basic inventive activity [may] exist without the grant of a patent”).

⁶⁰ Purists may object that open source licenses routinely assert copyright and are themselves a form of intellectual property. However, volunteers hardly ever participate because of copyright incentives. Instead, most traditional collaborations rely on softer incentives like altruism or a desire for education. Copyright’s role tends to be secondary and is usually involves deterring negative behaviors that might otherwise destabilize the collaboration. See S. Maurer, “The Penguin and the Cartel,” *Utah Law Review* (1): 262-318 (2012).

⁶¹ The idea makes an early appearance in H. G. Wells, *The World Set Free* (1914), <http://www.gutenberg.org/ebooks/1059>. Wells imagines a nuclear-powered world in which people respond to increased leisure time by becoming more creative. Predictably, the novel later takes a darker turn.

⁶² Wesley M. Cohen, Richard R. Nelson and John P. Walsh, “Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (Or Not),” National Bureau of Economic Research Working Paper 7552 (2004).

⁶³ *Graham v. John Deere*, 383 US 1 (1966)

⁶⁴ *Id* at 10 (quoting Thomas Jefferson).

dividing line where it did. To the contrary: Basing the standard on “average” skill sounds like a confession of ignorance. Why not choose “above-average” skill or a “blinding flash of inspiration?”⁶⁵ We discuss these and other alternatives in Section 5.

B. Goal No. 2: Selecting the Right Level of Effort.

An inventor can almost always create products sooner or more reliably if she is willing to spend more. What she actually does depends on the reward. Here the most natural guess is that the inventor will only invest where the expected reward⁶⁶ covers her R&D cost. On the other hand, the expected reward can be smaller if the inventor expects to supplement her income with non-patent income from prizes, first mover advantages, and the like. Conversely, society could rationally decide that it is worth higher prices to get inventions sooner or more reliably. This, too, would imply higher rewards for successful invention.

At the same time, bigger rewards imply broader and/or longer-lasting monopolies. Given diminishing returns, we should expect each increment of reward to provide less benefit than the one before. For this reason, theory predicts that there is some maximum patent duration and breadth beyond which further increases are counterproductive. In Prof. Kitch’s elegant summary, “The courts, influenced by the reward theory, view the patent system as a difficult problem of trade-offs between the incentive effects and the output constraining effects.”⁶⁷ Once again, the correct tradeoff is heavily fact-dependent so that theory provides only limited guidance.

Saying that a patent should be “narrow” or “broad,” of course, is not very precise. Doctrinally, the concept of breadth embraces (a) whatever the inventor actually claimed and disclosed in her patent, together with (b) certain “non-equivalent” improvements and follow-on inventions. While the latter doctrine is vague, we argue in Section 7 that it almost certainly implies a PHOSITA-like cleverness standard. In practice, judges and commentators usually assume that this test is identical to *Graham*’s non-obviousness bar. However, this is not the only choice and we will see that Congress could pick a different cleverness test if it wanted to.

⁶⁵ *Cuno Corp. v. Automatic Devices*, 314 US 84 (1941).

⁶⁶ The “expected value” of a reward refers to the average of all possible outcomes weighted by the probability that each outcome will actually happen. Wikipedia, “Expected Value,” http://en.wikipedia.org/wiki/Expected_value.

⁶⁷ Kitch, *supra*, at 282.

C. Goal No. 3: Allocating Reward Among Multiple Inventors.

We have already noted that many inventions are “cumulative” in the sense that progress depends on insights contributed by multiple, often-independent inventors over time. Consider, for the sake of definiteness, the case where a first generation (“1G”) invention establishes a new technology that receives 2G/3G/4G improvements and culminates in a saleable 5G product. Assume further that success is uncertain because the 5G product may turn out to be technically impossible or less popular than investors originally hoped. Then an economically efficient patent system must satisfy at least three separate criteria:

Allocate Profits. The 1G (2G, 3G...) R&D programs will never be funded unless patent owners expect to share in 5G revenues. Provided that the 5G product earns enough revenue, each inventor in the chain should expect to recover his or her investment.

Allocate Losses. The 1G (2G, 3G...) investors know that they could invest more on R&D than the 5G product will ever earn. In this case, someone must bear the loss. Here, the principle of “sunk costs” insists that investors at each stage should ignore “retrospective (past) costs that have already been incurred and cannot be recovered.”⁶⁸ Instead, they should base their decision exclusively on the costs of going forward.

Maximum Efficient Investment. Deciding to invest in an R&D program is not the only question; investors must also decide how much. In general, we expect each incremental increase in R&D budget to (on average) deliver more and better products to consumers. This investment should only be made, however, if these incremental benefits exceed incremental cost. This correctly prices the option value of new technologies.

The economics literature provides a clear rule that gets these incentives right.⁶⁹ Each inventor should (a) retain enough revenue to cover her R&D cost,⁷⁰ and (b) pay the balance to whichever inventor immediately preceded her. Following Prof. Tur-Sinai, we shall refer to this as the “Absolute Scope” rule in what follows.⁷¹

⁶⁸ Wikipedia, “Sunk Costs,” http://en.wikipedia.org/wiki/Sunk_costs.

⁶⁹ The proposed rule follows Scotchmer, *supra*. In principle, there may be other rules that satisfy our three criteria. If so, I have been unable to find them.

⁷⁰ More specifically, we want investors to recover revenue consistent with our first and second goals. Rewards that simply repay R&D investments may be insufficient to the extent that inventors lack, for example, access to capital.

⁷¹ Ofer Tur-Sinai, *supra*, 50 *IDEA* at 758 and note 142.

D. Reimagining Patent Doctrine.

The next six sections ask how well current patent rules conform to modern economic insights. We proceed as follows. Chapter 4 (“Reimagining Utility”) asks when an invention with no immediate end-use nevertheless offers sufficient social value to be patented. Chapter 5 (“Reimagining the PHOSITA”) argues that the *Graham* standard is best understood as a measure of cleverness and, indirectly, R&D cost. In the process, we generalize the PHOSITA concept to include the full range of cleverness standards available to Congress. Chapters 6 - 8 analyze how well patent doctrines that include the PHOSITA concept implement our third policy goal. Finally, Chapters 9 and 10 look at the current system’s shortcomings and ask whether radical change is worth pursuing.

4. Reimagining Utility

Any attempt to allocate reward among sequential inventors must start with the threshold question of when 1G (2G, 3G...) inventors should be allowed to share in 2G (3G, 4G...) revenues. Following the innovation economics literature, we argue that sharing should take place whenever inventions create “option value” for later inventors to build on.⁷² We begin by reviewing judges’ efforts to generalize the utility concept to inventions that have no present use and whose worth is based entirely on option value. We then suggest a revised rule that rationalizes earlier cases and brings doctrine into line with our third policy goal.

A. Modern Utility Doctrine.

Early US law provided that only “sufficiently useful and important” inventions could be patented.⁷³ This seemed to say that patent doctrine required applicants to show some minimum usefulness to consumers in the same way that *Graham*’s non-obviousness standard requires them to show cleverness.⁷⁴ But this did not last long. In 1817, Judge Story downgraded the utility requirement to “capable of use” and not “frivolous.”⁷⁵ This implied a *de minimis*

⁷² See Jerry R. Green and Suzanne Scotchmer (1995), *supra* note 43.

⁷³ Patent Act of 1836, ch. 357, § 7, 5 Stat. 117, 120 (1836) (repealed 1870).

⁷⁴ Michael Risch, “Reinventing Usefulness,” 2010 *BYU L Rev.* 1195, 1207, 1236 (2010). Some federal courts continued to describe non-*de minimis* utility thresholds in dictum well into the 20th Century. *Id.* The heresy was so prevalent that contemporary scholars and jurists repeatedly warned practitioners not to confuse “positive utility” with “degree.” N. Scott Pierce, “*In re Dane K. Fisher: An Exercise in Utility*,” 6 *J. High Tech. L.* 1 (2006); *Kirk* 376 F.2d at 995 (1967).

⁷⁵ *Lowell v. Lewis* 15 F. Cas. 1018, 1019 (CC Mass. 1817) (No. 8565); see also *Bedford v. Hunt*, 3 F. Cas. 37, 37 (CC Mass. 1817) (No. 1217) (invention must be “capable of use.”) The opinion famously added that inventions should

standard in which “any quantity of utility should suffice.”⁷⁶ Judge Story justified this result on the theory that patents on worthless inventions were irrelevant in any case, since such learning would “silently sink into contempt and disregard.”⁷⁷ In modern phrasing, the point seems to be that the damage from monopoly (*i.e.* deadweight loss) can never be larger than the monopoly itself. For this reason, there is nothing to be gained from refusing to patent inventions that nobody wants. We can also glimpse a second argument in Story’s comment: Given that patenting costs money, no rational inventor will ever pursue rights to an invention that he expects to be “disregarded.” In this situation, the inventor’s willingness to invest already provides a market test of value. A utility threshold that encourages judges to second-guess this judgment is, at best, superfluous and could easily end by overruling investments that the inventor sincerely believes in.

For the most part, Judge Story’s rules are still valid today. Starting around 1950, however, courts reintroduced a non-*de minimis* threshold for chemical and biology inventions.⁷⁸ The basic doctrinal puzzle was that these industries – unlike, say the mechanical or electrical arts – were willing to invest in “methods for producing intermediary products or products with unknown results.”⁷⁹ On the face of things, courts took a hard line by demanding that some use must “presently exist.”⁸⁰ This, however, would have ruled out intermediate inventions that commercial firms were plainly willing to pay for. For this reason, judges softened the rule by noting that utility could still be shown where persons “skilled in the art” would immediately recognize the product’s capabilities provided that they were not “conjecture” or else so

also not be “injurious to the well-being, good policy, or sound morals of society.” The Federal Circuit overruled this “moral utility” element in *Juicy Whip, Inc. v. Orange Bang, Inc.* 185 F.3d 1364 (Fed. Cir. 1999).

⁷⁶ Risch, *supra*, 210 *BYU L. Rev.* 1195. The invention did not have to perform better than competing prior art technologies. *In re Ratti*, 270 F.2d 810, 815 (C.C.P.A. 1959) (invention “no better than” the prior art was nevertheless patentable).

⁷⁷ *Lowell v. Lewis* 15 F. Cas. 1018, 1019 (CC Mass. 1817) (No. 8565).

⁷⁸ According to Judge Rich, pre-1950 chemical patent applications “were commonly granted although no resulting end use was stated or the statement was in extremely broad terms.” *In re Kirk*, 376 F.2d 936, 950 (1967) (Rich, dissenting). Judge Rich believed that the Story rule should apply to both intermediate and end-use discoveries. *Id.* at 1142.

⁷⁹ Michael John Gulliford, “Much Ado About Gene Patents: The Role of Foreseeability,” 34 *Seton Hall L. Rev.* 711, 719 (2004); *see also*, Tashica T. Williams, “Intellectual Property: Patent: *In re Fisher*: Raising the Utility Hurdle for Express Sequence Tags,” 21 *Berkeley Tech. L.J.* 123, 137 (2006) (creation of new chemical and biological compounds “frequently occurs long before the discovery of any specific utility”).

⁸⁰ *Ex parte Tolkmith* 102 USPQ 464, 466 (Pat. Off. Bd. App. 1954).

“unpredictable” as to require “considerable experimentation.”⁸¹ Even so, the rule ignored the fact that many industries really do fund 2G projects based on unproven conjectures. In these cases, at least, doctrine failed to protect option value.

The US Supreme Court tried to rationalize the situation in 1964. The *Brenner* opinion asked whether a chemical compound whose therapeutic effects had never been tested should nevertheless be patentable. The justices rejected this expansive result and instead announced that intermediate inventions would only qualify for protection if they possess “specific benefit” and “substantial utility” in a “currently available form.”⁸² The Court also advanced two policy arguments to justify its holding. First, it raised the familiar point that inventions should relate to “commerce” instead of “philosophy.”⁸³ This distinction, however, seemed redundant since the same concepts also helped to define patentable subject matter.⁸⁴ Just what did utility add to the exercise?⁸⁵ Perhaps sensing the problem, the Court introduced a second and more novel objection: “A patent,” Justice Harlan observed, should be more than a “hunting license.”⁸⁶ While this sounds like a rejection of option value, the Court was more likely referring to a very

⁸¹ *Id.* at 466. The fact that the compound was new did not bar patentability provided that it “belong[ed] to a class of compounds the members of which have become well recognized to be useful for a particular purpose, and it is evident that it is within the skill of the art to use the claimed compound for this purpose.” *Ex parte Ladd, Harvey, Cable and Szayna*, 112 U.S.P.Q. 337, 338 (Pat Off. Bd. App. 1955).

⁸² *Brenner v. Manson*, 383 U.S. 519 (1966), 534-535. For an extended exegesis of the case, see Donald L. Zuhn, Jr. “DNA Patentability: Shutting the Door to the Utility Requirement,” 34 *J. Marshall. L. Rev.* 973 (2001).

⁸³ *Id.* at 536. The court’s argument continues to influence current scholars who argue that utility provides a “timing” function for determining when an invention is “ripe.” Rebecca S. Eisenberg, “Analyze This: A Law and Economics Agenda for the Patent System,” 53 *Vand. L. Rev.* 2081, 2087 (2000); Rebecca S. Eisenberg and Robert P. Merges, “Opinion Letter as to the Patentability of Certain Inventions Associated With the Identification of Partial cDNA Sequences,” 23 *AIPLA Q.J.* 1, 6 (1995) (Utility “serves a timing function, leaving basic research discoveries in the public domain until they have yielded tangible benefits and have thereby left ‘the realm of philosophy’ and entered ‘the world of commerce.’”); Andrew T. Kight, “Pregnant with Ambiguity: Credibility and the PTO Utility Guidelines in Light of *Brenner*,” 73 *Ind. L.J.* 997, 1012 (1998) (“the utility requirement operates to distinguish between basic research and applied technology.”); Arti Rai, “Fostering Cumulative Innovation in the Biopharmaceutical Industry: The Role of Patents and Antitrust,” 16 *Berkeley Tech. L.J.* 813, 839 (2001) (“In practice, however, on the utility requirement serves as a particularly good proxy for differentiating upstream from downstream research”); Michael Risch, “Reinventing Usefulness,” 2010 *B.Y.U.L.* 1195, 1209 (2010) (“basic science, no matter how important and valuable, does not merit protection and is therefore not useful in the patent sense”).

⁸⁴ For the Supreme Court’s most recent attempt to explicate patentable subject matter, see, *Bilski v. Kappos*, 130 S. Ct. 3218, 561 US ___, 177 L. Ed. 2d 792 (2010),

⁸⁵ There is, of course, no rule against using multiple redundant doctrines to enforce a single policy goal. Indeed, this can be forced on judges by poorly drafted statutes or tangled precedent. At the same time, the practice invites confusion and risks “using up” levers that may be needed for other policy goals.

⁸⁶ *Brenner, supra*, at 536.

different problem in which applicants patent worthless inventions in hopes of capturing after-arising technologies that they did nothing to facilitate.⁸⁷

Courts have now spent half a century trying to generalize *Brenner* beyond its facts. The earliest cases staked out an obvious bright line rule: The utility threshold excludes products whose only known use is to make other products that similarly lack known uses.⁸⁸ This rule was most recently applied in *In Re Fisher*, which denied patentability to a list of genetic sequences – so-called “Expressed Sequence Tags” or “ESTs” – that lack known functions. The reason, according to the Federal Circuit, is that ESTs are mere “research intermediates” presenting no “assurance that anything useful will be discovered in the end.”⁸⁹

The case is harder for inventions that claim specific uses. Here, post-*Brenner* utility analyses usually devolved into arguments over whether the evidence was sufficient to prove the asserted capability. Significantly, the required proof was much lower than the human trials that were needed to obtain FDA approval and achieve “commercial usefulness.”⁹⁰ Instead, utility could be established by animal tests and chemical similarity to previously-known drugs⁹¹ so long as these methods offered a “reasonable correlation” or were “generally predictive” of end-

⁸⁷ Because all legal systems are inherently imperfect, patent owners can sometimes assert accidents of language to claim inventions that were actually developed by others. Judges have long resisted this outcome. Prof. Zuhn points out that common law mining law similarly does not let miners claim whatever gold deposits lie beneath randomly selected plots of land. Donald L. Zuhn, Jr. “DNA Patentability: Shutting the Door to the Utility Requirement,” 34 *J. Marshall. L. Rev.* 973 (2001). See also, William E. Ridgway, “Realizing Two-Tiered Innovation Policy Through Drug Regulation,” 58 *Stan. L. Rev.* 1221 (2006) (setting the utility threshold too low leads to inventors “stockpiling” instead of “developing” innovations.)

⁸⁸ *In re Kirk*, 376 F.2d 936, 945 (1967)(“it is not enough that the specification disclose that the intermediate exists and that it ‘works,’ reacts, or can be used to produce some intended product of no known use.”); *In re Joly* 376 F.2d 906, 908 (utility was not demonstrated by “mere disclosure that a claimed chemical compound can be used as an intermediate to make other compounds, without regard for the usefulness of the latter compounds”). Significantly, the court’s holdings were based on the fact that many steroids “possess no activity whatsoever” and that the patented compound’s “actual uses...or possible lack of uses” could only be determined by additional experimentation. *Kirk* at 1125-26. These uncertainties did not, of course, show that the discoveries lacked sufficient “option value” to justify 2G research.

⁸⁹ *In re Fisher*, 421 F.3d 1365 (2005). Judge Rader’s exasperated dissent cogently argued that any research which took society “one step closer to identifying and understanding a previously unknown and invisible structure” had social value and that this value existed whether or not the particular invention represented “the final step of a lengthy incremental research inquiry...” *Id.* at 1380 (Rader dissent).

⁹⁰ *In re Langer*, 503 F.2d 1380, 1393 (CCPA 1974); *Accord, Janssen Pharmaceutica v. Teva Pharmaceuticals USA, Inc.*, 583 F.3d 1317, 1324-25 (Fed. Cir. 2009) (same). This made some sense. Absent FDA regulation, adventurous (or desperate) consumers might well purchase drugs based on animal test results and/or a molecular resemblance to known compounds.

⁹¹ *In re Jolles*, 628 F.2d 1322 (1980).

use capabilities.⁹² Not surprisingly, judges identified such proofs with the PHOSITA concept, *i.e.* whether “one of ordinary skill in the art would accept appellant’s claimed utility in humans as valid and correct”⁹³ or else “would not doubt the utility” based on animal testing.⁹⁴ While this lent the analysis an aura of scientific objectivity, it also begged the question of how a PHOSITA would decide just when the numeric odds of success were large enough to be deemed “reasonable” or “generally predictive.” But in that case any reasonable answer would have involved some judgment about which 2G projects are worth pursuing – an inquiry that necessarily includes non-technical factors like researcher salaries and consumer demand. Not for the first time, the PHOSITA standard let judges sweep these economic variables under the rug.⁹⁵

Despite doctrinal gymnastics, then, courts continue to tie utility to presently-existing capabilities that – however hypothetically – consumers might actually want to use. Strangely, however, judicial policy arguments are seldom so narrow. For example, the Court of Customs and Patent Appeals argued in *Cross* that a low utility threshold is needed to “marshal resources and direct the expenditure of effort to further *in vivo* testing...thereby providing an immediate benefit to the public.”⁹⁶ Similarly, the Federal Circuit has held that *Brenner* utility “necessarily includes the expectation of further research and development.”⁹⁷ These arguments appear indistinguishable from option value.

B. A Rationalized Utility Standard

The traditional utility test for ordinary “present value” inventions is simple: Does the invention do something that consumers might value? The problem, as we have seen, is that industries often find commercial reasons to invest in inventions that have no known end-use and hence

⁹² *Cross v. Iizuka*, 753 F.2d 1040 (Fed. Cir. 1985)

⁹³ *In re Jolles*, *supra*, at 1327-28, see also US Patent and Trademark Office, Utility Examination Guidelines, 66 Fed. Reg. 1092, 1093 (Jan 5 2001) (arguing that genes should be patentable where “a person of ordinary skill in the art would immediately appreciate why the invention is useful.”); *In re Tolkmith*, 102 USPQ 464, 465, 466 (arguing that value of proposed parasiticide is measured by whether its capabilities are apparent to “one skilled in the art.”)

⁹⁴ *In re Brana*, 51 F.3d 1560 (Fed. Cir. 1995). *But see, ex parte Aggarwal*, 23 USPQ2d 1334 (1992) (evidence that compounds treated tumors in animals did not provide sufficient evidentiary support for human applications.)

⁹⁵ Courts also debate how “specific” the claimed end use has to be. While the requirement is clear enough for compounds that treat specific diseases, judges have stretched it to include “any compound” that demonstrates “known pharmacological activities” like reducing blood pressure. *Nelson v. Bowler*, 626 F.2d 853, 855 - 856 (CCPA 1980).

⁹⁶ *Cross*, *supra* at 1051.

⁹⁷ *In re Brana*, *supra*, 51 F.3d 1560 at 1568.

no value to consumers. This puts judges and bureaucrats in the awkward position of telling industry that plainly commercial R&D lacks value “in the real world.”⁹⁸

Following Judge Story, we argue that it is simpler – and also more economically efficient – to define value with reference to actual consumers and markets. The question is, *which* markets and *which* consumers? Current doctrine invariably identifies “markets” and “consumers” with end-users. This, however, requires fictitious inquiries into how much consumers would value inventions that will never be sold to the public. But this fiction is unnecessary: After all, there really is a market for 1G results. Following the innovation literature, we argue that the real consumers are 2G inventors. More specifically, we say that a 1G invention only has value if someone, somewhere is willing to develop it.⁹⁹ This transforms the issue of what end-users might or might not value into an objective investment decision. The rule is also consistent. Inventions that fail to attract 2G investors are automatically worthless to follow-on (3G, 4G, ...) developers, end-users, and society at large.

This option value viewpoint is already implicit in judicial statements tests that measure value from the PHOSITA’s standpoint. However, economic theory considerably sharpens this instinct. Under our rationalized rule, an intermediate invention has value any time an inventor would pay a non-zero (but possibly *de minimis*) fee to develop it. Assuming profit-maximization,¹⁰⁰ this judgment normally turns on three well-defined variables:

Probability of Success. The 2G program must have some chance of success. We have already noted that current patent doctrine stresses this scientific judgment.

Expected R&D Cost. The 2G project must not be “too costly” to develop. Naively, this will be true whenever the expected reward is even slightly greater than expected costs. This neatly replicates Judge Story’s *de minimis* rule that inventions should be patentable whenever their economic value is even slightly greater than zero.

⁹⁸ Compounds used solely for industrial R&D are said to lack “real world” value as a matter of law. See, e.g., U.S. Patent & Trademark Office, *Revised Interim Utility Guidelines Training Materials* 5, <http://www.uspto.gov/web/menu/utility.pdf>.

⁹⁹ In the case of research tools, one might equally say “use.” We ignore this essentially semantic distinction in what follows.

¹⁰⁰ Our rule does not really need this assumption. The chance that a 1G invention may lead to further development by non-commercial actors like universities, national laboratories, and open source collaborations is also included in option value.

Expected Revenue. The 2G decision ultimately depends on a judgment that an end product would earn sufficient revenue to cover R&D cost.¹⁰¹ This is also an economic judgment.

The nice thing about using a PHOSITA standard is that it avoids making these inquiries directly. Instead, following Judge Story, the PHOSITA's judgment that further investment is worthwhile serves as a conclusive test of value. While litigants can address our three economic variables directly, it will often be enough to submit less direct evidence. This could include the opinions of knowledgeable industry observers or objective evidence that similar ideas have been funded in the past. At the same time, our explicitly economic rule reminds courts that such showings can be rebutted by evidence that option value has changed over time. This notably includes cases where R&D costs fall or consumer demand increases. Finally, courts should recognize that option value can sometimes exist even where the 1G inventor cannot name a single 2G application.¹⁰² This can happen where the 1G inventor makes a fundamental discovery but knows almost nothing about other industries or else reasonably believes that some PHOSITA-95 inventor will find applications she never thought of.

5. Reimagining the PHOSITA

Economists and lawyers speak different languages. This naturally leads to problems. We have seen that each of our three economic principles is defined by R&D cost. Yet patent law hardly even mentions this concept.¹⁰³ This is not necessarily fatal since law often uses proxies for quantities that are difficult or impossible to measure directly. But if a proxy does exist, we should be able to name it. Fortunately, the choices are limited. Open any introductory textbook and you will see that patent law is constructed around a dozen or so core concepts.¹⁰⁴ Of these, only two – “non-obviousness” and “utility” – say anything about what the inventor has

¹⁰¹ The judgment may, of course, depend recursively on a perception that 3G inventors are willing to pay licensing fees. But in that case the 3G inventor must believe that a 4G inventor would pay, and so on. Ultimately, the chain ultimately depends on a revenue estimate.

¹⁰² The best-known example is almost certainly the laser, which many contemporaries saw as a “solution looking for a problem.” Charles H. Townes, “The First Laser,” in Laura Garwin and Tim Lincoln (eds.) *A Century of Nature: Twenty-One Discoveries that Changed Science and the World* (Chicago U. Press 2003) at p. 107.

¹⁰³ As we shall see, the concept of R&D cost plays little or no role in the utility (Section 7), non-obviousness (Section 5B), or blocking patents (Section 8B) doctrines. It also plays no role in damages: None of the fifteen *Georgia Pacific* factors so much as mentions it. *Georgia-Pacific Corp. v. United States Plywood Corp.*, 318 F. Supp. 1116, 6 U.S.P.Q. 235 (S.D.N.Y. 1970).

¹⁰⁴ See, e.g. Robert Merges, Peter Menell and Mark Lemley, *Intellectual Property in the New Technological Age* (5th ed. Aspen: 2010) at pp. xii – xv.

accomplished or, implicitly, invested. Furthermore, we have already seen that utility measures value from the consumer's perspective. This is a very different concept from the inventor's effort or cost. In the end, we are left with a single candidate proxy: Non-obviousness.¹⁰⁵

This section introduces *Graham's* PHOSITA test and explains how the idea can be generalized to measure different levels of cleverness and, indirectly, effort and cost. We then examine the PHOSITA test's strengths and weaknesses as a cost-proxy. We close by commenting on how the proxy limits doctrine's ability to implement economically efficient incentives.

A. The PHOSITA Standard

The Federal Circuit's efforts to make patent law more rule-based and predictable have encrusted the PHOSITA with properties that no actual person possesses. On the one hand, she is a polymath who possesses "a complete and thorough knowledge of all legally pertinent prior art." On the other, she is an uninspired drudge who thinks "almost entirely along the line of conventional wisdom" and brings little or no originality to R&D.¹⁰⁶ Clearly, we should not expect to meet this person in real life. Like the proverbial "reasonable man," she is best seen as a construct or "ghost of the law."¹⁰⁷ We return to these points below.

For now, we take a different path. The phrase "ordinary skill" can be readily identified with "average" or perhaps "median" skill.¹⁰⁸ Compared to our judge-made "ghost of the law," these are objective and indeed measurable characteristics. In what follows, we will think of our

¹⁰⁵ We find further encouragement in the fact that judges tend to invoke the PHOSITA concept in the same contexts where innovation economists stress R&D cost. This includes reward size, non-obviousness, and even utility. See Sections 6B, 6C and 7, below.

¹⁰⁶ As Profs. Abramowitz and Duffy write, "the courts have defined the person of ordinary skill to be a rather extraordinary creature, an idiot savant with extraordinary knowledge and virtually no creativity. The mind of this hypothetical person comes equipped with a complete and thorough knowledge of all legally pertinent prior art, far more knowledge than could be possessed by any average or actual researcher.... Meanwhile, the Federal Circuit at one time described the person of ordinary skill as someone 'who thinks along the line of conventional wisdom in the art and is not one who undertakes to innovate.'" Michael Abramowitz & John Duffy, "The Inducement Standard of Patentability," 120 *Yale L.J.* 1590, 1606 (2011).

¹⁰⁷ *Panduit Corp. v. Dennison Mfg., Co.*, 810 F.2d 1561 at 1566 ("The decisionmaker confronts a ghost...not unlike the 'reasonable man' and other ghosts in the law."); see also, *Kimberly-Clark Corp. v. Johnson & Johnson* 745 F.2d 1473 (Fed. Cir. 1984) (PHOSITA is a "hypothetical person" and "an imaginary being" of the courts' "own devising").

¹⁰⁸ See, e.g., Dan L. Burk and Mark A. Lemley, "Biotechnology's Uncertainty Principle," 54 *Case W. Res.* 691, 712 (2004) ("The PHOSITA standard contemplates some median or ordinary skill.")

PHOSITA as a real person who has a name, a home address, pays taxes, and goes to work most mornings. If we wanted to, we could find her and put her on the witness stand.¹⁰⁹

We can also generalize the concept. As Prof. Scotchmer points out, there is no economic reason why the same PHOSITA concept should be used to define *both* non-obviousness *and* patent breadth. Indeed, de-linking the tests would solve various policy problems.¹¹⁰ Identifying the PHOSITA with real people makes this surprisingly easy to do. Recall that *Graham* defined non-obviousness in terms of a PHOSITA having “average skill in the art,” *i.e.* an invention that is cleverer than fifty percent of all R&D projects.¹¹¹ In what follows we will call this threshold “PHOSITA-50” where “50” denotes the invention’s percentage cleverness rank compared to all other inventions in the art. But in that case, why not generalize the concept to define additional thresholds for, say, “PHOSITA-33” or “PHOSITA-95”? Obviously, the number of available tests is infinite. We will use the term “PHOSITA-X” to denote this family of standards.¹¹²

B. How Good is the Proxy?

The PHOSITA-X concept lets juries determine cleverness from the face of the patent. This much seems obvious. The deeper question is whether cleverness is itself a proxy for R&D cost. We argue that there are at least four good reasons to link cleverness with cost:

Brute Force Research. “Genius,” as Thomas Edison remarked, is “one percent inspiration” and “ninety-nine percent perspiration.”¹¹³ He should know. Edison’s search

¹⁰⁹ Courts sometimes recognize this possibility. *Id.* at 710 (“Courts have on occasion equated the knowledge of a given individual, such as a patent examiner, with that of the PHOSITA.”)

¹¹⁰ Scotchmer, *supra* at 146-149; *see also* Section 7D below. Profs. Burk and Lemley note that it is doctrinally possible to use different PHOSITA standards in different legal contexts. Dan L. Burk and Mark A. Lemley, *supra*, 54 *Case W. Res.* at 712 (“The PHOSITA for purposes of obviousness may not necessarily be the PHOSITA for purposes of enablement, written description, definiteness or equivalence.”)

¹¹¹ Remarkably, our PHOSITA standard has a recursive quality: The cleverness it prescribes cannot be specified in isolation and differs from market to market. To see this, suppose that a PHOSITA-50 standard is implemented on Day One. Rational inventors whose inventions fall just below the threshold will immediately invest in making their products cleverer since patentability promises a large jump in value for very little effort. Now, however, this new investment raises the average cleverness level so that some inventions that satisfied the PHOSITA-50 standard on Day One no longer qualify. This encourages a second round of Day Two investments which leads to further Day Three investments and so on. The net result is that the PHOSITA standard improves the average cleverness of all inventions, including many that are never patented.

¹¹² As before, we use “X” to denote the invention’s percentage cleverness rank compared to all other inventions in the art.

¹¹³ Wikiquote, “Thomas Edison.” http://en.wikiquote.org/wiki/Thomas_Edison.

for a workable light bulb required tests of more than 3,000 filaments.¹¹⁴ While cleverness may have saved him from testing still more candidates, Edison himself was doubtful.¹¹⁵ Similar brute force searches are common in many modern industries including pharmaceuticals.¹¹⁶

Corporate Research. Team research is inevitable in any industry – for example, airplanes and computer software – where innovation requires more man-hours than any human lifetime can supply. The fact that companies routinely hire R&D teams shows that cleverness really can be bought and sold as a commodity.

Self-Employed Inventors. Many inventors who work at home could be earning wages in commercial R&D projects. Given this opportunity cost, even garage-based R&D programs have an implicit dollar value.

New Ideas. The link between money and innovation is thinnest when we ask how innovators get R&D ideas in the first place. Even so, Pasteur reminds us that “Chance favors the prepared mind.”¹¹⁷ This suggests that paying people to think about a problem really can increase the supply of “prepared minds” and, implicitly, the flow of ideas.

These observations confirm the intuition that money can produce cleverness. Before using cleverness as a proxy, however, we also want to know how often the link is broken. After all, some valuable ideas really do cost nothing. In these cases, our proxy will overprice invention. This is a fundamental problem, though not nearly as bad as under-rewarding ideas so that innovation never occurs at all.

¹¹⁴ *The Incandescent Lamp Patent*, 159 US 465, 16 S. Ct. 75; 40 L. Ed. 221 (1895).

¹¹⁵ While Edison’s selections were theory-driven, the theories themselves were unreliable. As Edison later recalled, “I would construct a theory and work on its lines until I found it was untenable... I speak without exaggeration when I say that I have constructed 3,000 different theories in connection with the electric light, each one of them reasonable and apparently likely to be true. Yet only in two cases did my experiments prove the truth of my theory.” Wikiquote, “Thomas Edison,” *supra*.

¹¹⁶ Pharmaceutical companies test roughly 5,000 candidates for each drug approved. See, e.g. Cath Driscoll, A Virtual Space Odyssey, http://www.nature.com/horizon/chemicalspace/background/figs/odyssey_f1.html; see, also Asterand Corp., “About Us” (estimating 5,000-10,000 candidates per successful drug) <http://www.asterand.com/Asterand/about/index.htm>. Readers can find an extended discussion of the drug discovery pipeline in Solomon Nwaka and Robert G. Ridley “Virtual Drug Discovery and Development for Neglected Diseases Through Public–Private Partnerships,” *Nature Reviews: Drug Discovery* 2:919-928 (2003).

¹¹⁷ Wikiquote, “Louis Pasteur,” http://en.wikiquote.org/wiki/Louis_Pasteur.

Limitations. Using cleverness as a cost proxy has at least three drawbacks. First, R&D costs almost always vary from industry to industry. This means that the PHOSITA standard cannot be used to compare costs across industries. Patent law has long acknowledged this limitation.¹¹⁸ Second, the standard can also vary for different types of research within the same industry. This problem is particularly noticeable for pharmaceuticals, where costs usually increase several-fold at each step along the R&D “pipeline.”¹¹⁹ Third, the PHOSITA test only provides information on *relative* costs. While we know that some projects are costlier than others, we cannot say by how much. Finally, R&D costs are unpredictable. This suggests that the *ex post* cost of R&D programs may often be larger or smaller than inventors anticipate. This should not matter in most cases. This is because innovation economics almost always depends on *ex ante* incentives.

D. Uncertain Outcomes

Identifying the PHOSITA with cost is only half the battle. An economically efficient patent law must also be flexible. In order to implement our Absolute Scope rule, courts must be able to allocate revenue in all possible ratios from 100:0 to 0:100. On the face of things, however, patent doctrine offers just four possible outcomes: 0:0 (both patents invalid), 50:50 (both patents upheld), 100:0 (first patent valid, second invalid), or 0:100 (first patent invalid, second valid). Like the proverbial stopped clock, it seems that doctrine can only be right by accident.

Fortunately, closer inspection reveals a loophole. As Profs. Lemley and Shapiro have emphasized, real inventors make their investment decisions *before* litigation based on “expected value,” *i.e.* the average of all possible outcomes weighted by the probability that each outcome will actually happen.¹²⁰ In theory, at least, these probabilities can take on any value from 1 to 100. This means that suitably chosen probabilities can generate whatever allocations our Absolute Scope rule requires. From an *ex ante* incentives standpoint, patent law is more flexible than it looks.

¹¹⁸ Dan L. Burk and Mark A. Lemley, “Policy Levers in Patent Law,” 89 Va. L. Rev. 1575 (2003) at 1650 (“As the name suggests, PHOSITA-based analysis is specific to the particular art in which the invention is made. Courts measure most significant patent law doctrines against a benchmark that varies by industry....Overwhelming evidence indicates that the application of the PHOSITA standard varies by industry, leading for example to fewer, but broader, software patents and more, but narrower, biotechnology patents.”)

¹¹⁹ Published estimates for each stage of the process are collected in S. Maurer, “The Right Tool(s): Designing Cost-Effective Strategies for Neglected Disease Research,” paper commissioned by World Health Organization (2005). See also, Driscoll, A Virtual Space Odyssey, *supra*.

¹²⁰ Mark A. Lemley and Carl Shapiro, “Probabilistic Patents,” J. Econ. Perspectives 19(2):75-98 (2005). Profs. Green and Scotchmer also examine models in which patent validity is probabilistic. The main difference is that their model focuses on uncertainty in R&D outcomes rather than litigation. Jerry R. Green and Suzanne Scotchmer, “On the Division of Profit in Sequential Innovation,” RAND J. Econ. (26):20-33 (1995).

How patent law uses this flexibility is another matter. Plainly, the actual outcomes depend on doctrine. In Section 6, we show how doctrine translates the inherent uncertainty in jurors' PHOSITA-X determinations into expected allocation outcomes. While these allocations fail to replicate our Absolute Scope rule, they are nevertheless surprisingly reasonable.

E. Lessons for Doctrine.

Any doctrine that uses the PHOSITA concept as a cost proxy is bound to make errors. These errors can be roughly divided into (a) the way the concept is defined, and (b) its lack of predictability from one jury to the next. These simple observations already provide useful insights for doctrine.

Definitional Limits. There are many things that the PHOSITA concept cannot do even in principle. We have argued that sequential innovation is most efficient under an Absolute Scope rule where each inventor (a) retains enough revenue to cover her actual R&D cost, and then (b) pays any excess to her immediate predecessor. But a jury's PHOSITA-X determination does not measure actual R&D cost. It only provides a relative judgment that some inventions are cleverer – and, implicitly, was more expensive to develop – than others. *It follows that no PHOSITA-based doctrine can say when a particular inventor has covered her costs and should transfer any remaining revenue upstream.* This limitation deters investment by increasing the odds that at least one inventor will fail to recover her costs. For this reason, PHOSITA-based incentives can never be as efficient as our ideal Absolute Scope rule.

Uncertainty. Lawyers are trained to seek clear and replicable rules. However, we have argued that the PHOSITA proxy works best when jury determinations are uncertain. This is obviously troubling. Still, we should also be realistic. The Federal Circuit has so far tried – and failed – to make patent outcomes more predictable for the past thirty years.¹²¹ In these circumstances, it may be wiser to admit that the problem exists and domesticate its good features.

Fortunately, the economic efficiency costs are limited. If investors are risk-neutral, average expected reward is the only thing that matters. If they are not, they can readily manage the problem by investing in so many companies (and implicitly, patent portfolios) that any fluctuations cancel.

¹²¹ As Judges Mayer and Newman have protested in the context of claim construction, “In the name of uniformity, [we have] held that claim construction does not involve subsidiary or underlying questions of fact...What we have wrought, instead, is the substitution of a black box, as it so pejoratively has been said of the jury, with the black hole of this court.” *Phillips v. AWH Corp.*, 415 F.3d 1303, 1330 (2005).

The main costs of uncertainty involve individual justice, *i.e.* the certainty that some inventors will inevitably receive windfalls while others are penalized. Still, even this objection is weaker than it sounds. First, everyone knows that lawsuits are uncertain and patent owners necessarily accept this risk. Second, we have already said that investors can defend themselves by diversifying. Finally, the actual number of cases that go to trial in any well-functioning system is small. This means that most rewards will be paid in the form of licensing fees rather than court judgments. Theory suggests that most of these payouts will cluster somewhere near the average litigated outcome.¹²²

Reforming Doctrine. This Section began by remarking that the PHOSITA doctrine is considerably more complicated than our simple PHOSITA-X discussion suggests. Some of these additional rules are designed to make the jury inquiry more routinized and predictable¹²³ and would presumably improve our proxy. However, we have also seen that the PHOSITA doctrine has become encrusted with “reasonable man”-type requirements that have nothing to do with actual inventors. Judges who take our arguments seriously should want to clear away this underbrush.

6. How Doctrine Allocates Reward (A): Literal Infringement and Blocking Patents

We have argued that the PHOSITA concept supplies a reasonable proxy for R&D costs. However, doctrine must still provide rules that translate jurors’ PHOSITA determinations into expected rewards for 1G and 2G inventors. US patent law currently does this through a two-step process. First, the jury must find that the 1G patent is valid, *i.e.* satisfies *Graham’s* “non-obviousness” test. Second, the jury must find that the 2G invention infringes, *i.e.* falls within the 1G patent’s “breadth.” Strangely, this second inquiry is governed by two separate sets of rules – “Literal Infringement” and the “Doctrine of Equivalents.” This Section analyzes the simplest, “Literal Infringement” case. This typically occurs where the 1G patent literally describes all or part of the 2G invention. We discuss more complex doctrinal problems involving non-literal infringement and affirmative defenses in Sections 7 and 8.

¹²² Theorists traditionally assume that parties who know the likely litigation outcome will settle to avoid unnecessary transactions costs. The range of bargained outcomes depends on the size of those costs. The theory is potentially testable since it predicts that the parties will only litigate close questions. George Priest and Benjamin Klein, “The Selection of Disputes for Litigation,” 13 *J. Legal Studies* 1 (1984). Unfortunately, the prediction is highly fact-dependent and may not hold in practice. *See, e.g.,* Keith Hylton, Asymmetric Information and the Selection of Disputes for Litigation,” 22 *J. Legal Studies* 187 (1993); Steven Shavell, “Any Frequency of Plaintiff Victory at Trial is Possible,” 25 *J. Legal Stud.* 493 (1996).

¹²³ For a comprehensive survey, see *Phillips v. AWH Corp.*, 415 F.3d 1303 (Fed. Cir. 2005) (en banc).

This Section shows how the Literal Infringement doctrine uses the PHOSITA concept to allocate reward between successive inventors. Section A briefly describes the Literal Infringement doctrine and related legal rules for allocating reward between 1G and 2G inventors. Section B asks how these formal rules translate into the *ex ante* incentives to invest in 1G and 2G inventions. Section C briefly reviews what is known about patent litigation uncertainty. Section D comments on our analysis's implications for doctrine.

A. Literal Infringement and Blocking Patents

The boundary between Literal Infringement and the Doctrine of Equivalents runs like a fissure through patent doctrine.¹²⁴ As the Supreme Court's *Graver Tank* decision explains, "resort must be had in the first instance to the words of the claim. If accused matter falls clearly within the claim, infringement is made out and that is the end of it."¹²⁵ This necessarily excludes all evidence of the 2G invention's cleverness.¹²⁶

Literal Infringement frequently leads to "Blocking Patent" scenarios in which anyone practicing the 2G invention must obtain licenses from *both* the 1G *and* 2G patent owners. Formally, Blocking Patents occur whenever both inventions are patentable and the second patent "includes the first."¹²⁷ This usually involves situations where the 1G and 2G inventions are both patentable and a complete description of the 2G invention includes (a) all of the underlying 1G invention's original claims elements, and (b) additional claims elements supplied by the 2G inventor.¹²⁸ Conversely, 2G inventions that modify one or more elements of the 1G patent escape literal infringement and are analyzed under the Doctrine of Equivalents.

¹²⁴ See generally, *Chisum on Patents*, § 18.04[4][a] (Federal Circuit decisions "regularly refer to 'literal infringement' as one of the two species of infringement, the other being infringement under the doctrine of equivalents") and cases cited therein.

¹²⁵ *Graver Tank & Mfg. Co. v. Linde Air Prod. Co.*, 339 US 605, 607 (1950).

¹²⁶ For example, evidence showing that the 2G invention has received a patent is inadmissible in literal infringement cases. See, e.g., *Chisum on Patents* at 16-21 and cases cited therein.

¹²⁷ *Cantrell v. Wallick* (1886) 117 US 689.

¹²⁸ *Atlas Powder Co. v. E.I. Du Pont de Nemours & Co.*, 750 F.2d 1569, 1580 (Fed. Cir. 1984) ("[I]f Atlas patents A + B + C and Du Pont then patents the improvement A + B + C + D, Du Pont is liable to Atlas for any manufacture, use, or sale of A + B + C."); see also, Donald Chisum, "The Scope of Protection for Patents after the Supreme Court's Warner-Jenkinson Decision: Fair Protection – Certainty Conundrum," 14 *Santa Clara Computer & High Tech. L.J.* 1 (1998) ("The patentability of an accused product or process often stems from what may be characterized as an additive or selective improvement, adding elements, features, or functions to the earlier patented invention or discovering preferred species for generic elements in the earlier invention, rather than from a mere substitution of one element for another. In such cases, Federal Circuit

B. What Incentives Do Investors See?

We now ask how the *Graham* standard allocates patent revenue between successive innovators. This Section introduces the basic logic by analyzing a simple scenario in which a 2G inventor patents improvements to a previously patented 1G technology. For simplicity, we assume that the 1G product is pre-commercial and earns no revenue. The 1G inventor sues for infringement and each party argues that the other's patent is invalid under *Graham*'s non-obviousness test.

Analysis. We begin by reviewing the possible legal outcomes. The *Graham* "non-obviousness" test tells juries to (a) determine each invention's PHOSITA-X rank, and (b) invalidate patents that do not exceed PHOSITA-50. Since invalid patents receive nothing, our hypothetical lawsuit can end in 0:0, 100:0, and 0:100 allocations. If both patents are upheld, our blocking patents assumption produces 50:50 sharing.

In order to decide the case, the jury must determine PHOSITA-X values for both the 1G and 2G patents. We start with the 1G patent. In theory, at least, we can imagine empaneling 1,000 different juries to determine the 1G patent's PHOSITA-X value. Suppose that we do this and the resulting verdicts have, say, a mean of 55. Since all 1,000 juries have seen the same evidence and received the same instructions, it is reasonable to think that most juries will reach similar results. This suggests that juries will pick rankings close to 55 (*e.g.* 58) more often than distant ones (*e.g.* 75). The solid line in Fig. 1 depicts a typical example of this "verdict distribution." Loosely speaking, readers can think of the vertical axis as showing the probability that a jury will

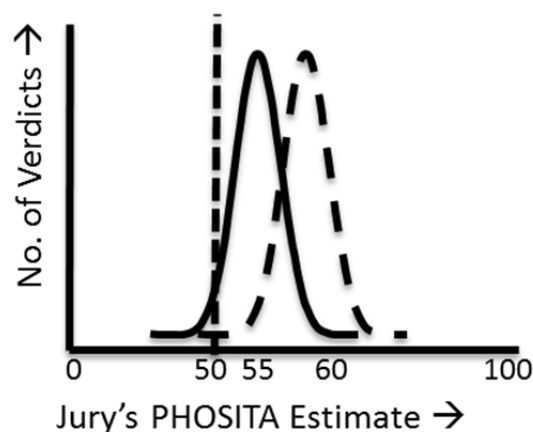


Fig. 1

decisions recognize that an accused product or process does not avoid infringement by adding functions or features if it contains literally or by equivalents all the elements of patent claim in question.")

determine each rank from 1 to 100.¹²⁹

Different inventions will usually have different verdict distributions.¹³⁰ How does this affect the allocation of reward? We know that the expected allocation is determined the ratio of how often the 1G invention is upheld to how often the 2G invention is upheld. Geometrically, this corresponds to how much of the area under each verdict distribution falls to the right of *Graham*'s PHOSITA-50 standard. On the other hand, a PHOSITA-60 invention (Fig. 1, dashed line) will usually have more area to the right of the PHOSITA-50 threshold than a PHOSITA-55 invention (Fig. 1, solid line). More clever inventions will therefore be upheld more often – and have greater expected value – than less clever ones. This is encouraging.

As usual, the devil is in the details. Ideally, we would like to use our verdict distributions to allocate reward according to cost.¹³¹ But this requires proportionality: A one percent increase in cleverness should improve an invention's chances of receiving revenue by roughly one percent. In practice there are distortions. A glance at Fig. 1 shows why. Our example started with a verdict distribution that peaked near *Graham*'s PHOSITA-50 threshold standard. Because the distribution is so tall at this point, even a slight increase in cleverness drags a large slice of area into the right hand side. In this situation, the patent system tends to over-reward small cleverness increments. On the other hand, suppose that distribution's peak is far to the right of the threshold. Here, a one percent increase in cleverness generates too little expected reward. Fig. 2 depicts the extreme case where both inventions are extremely clever [Fig.2]. Here, the chances that jury's PHOSITA-X determination will be less than 50 are already tiny. For this reason, further increases in cleverness have almost no impact on expected revenue. Instead, the allocation is dictated by our 50:50 blocking patents default rule.

¹²⁹ Mathematically-inclined readers will recognize that one cannot really speak of "the probability of x" for continuous values. Properly speaking, we should employ "probability densities" instead.

¹³⁰ If our legal rules were completely determinate, all 1,000 juries would return a verdict of 55. In this case Fig. 1 would show an infinitely narrow verdict distribution, *i.e.* a vertical line. More realistically, we expect real verdicts to contain a random component as well. Where this component is small, we expect juries to deliver "mostly determinate" verdicts. This will produce narrow verdict distributions with high peaks. Where the random component is larger, we expect verdict distributions to be broader and flatter.

¹³¹ We have already said that this prescription follows the neoclassical economics literature. It is worth noting, however, that some legal scholars suggest that reward should reflect value to consumers. *See, e.g.* Amy L. Landers, "Patent Claim Apportionment, Patentee Injury, and Sequential Innovation," 19 *Geo. Mason L. Rev.* 471 (2012). From this standpoint, the current system of tying reward allocation to the PHOSITA standard is a disaster. Perhaps the most natural solution would be to rephrase *Graham*'s PHOSITA standard to require above-average utility instead. US law has repeatedly rejected this innovation. Michael Risch, "Reinventing Usefulness," 2010 *BYU L Rev.* 1195, 1207, 1236 (2010)

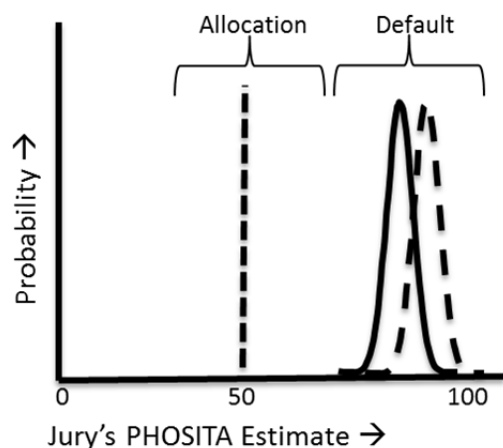


Fig. 2

For convenience, we will refer to regions close to the PHOSITA-50 as the “Allocation Zone” in what follows. Conversely, we will refer to regions dominated by the 50:50 rule as the “Default Zone.”

Putting these results together, we distinguish several cases. First, both patents may fall inside the Allocation Zone. Here, the expected allocations really do depend on comparative cleverness. Second, both patents may fall far to the right of the PHOSITA-50 threshold. Within this Default Zone, we expect a blocking patents rule to enforce 50:50 sharing even when one patent is significantly cleverer than the other. Finally, readers should consider the case (not shown) where one patent falls in the Default zone and the other falls inside the Allocation zone. In this case, the final result will depend almost entirely on small differences in the Allocation Zone patent’s cleverness while ignoring similar nuances in the Default Zone patent.

Finally, we have drawn Figs. 1 and 2 so that they feature narrow verdict distributions with high, sharp peaks. For large uncertainties, the verdict distributions become so broad that the Default Zone disappears entirely.

D. Which World Do We Live In?

Our argument suggests that – far from being a flaw – litigation uncertainty provides an essential bridge between patent law’s formal winner-take-all outcomes and the need for practical incentives that promise adequate reward to each inventor in the R&D chain. Clearly, much depends on the shape of our verdict distributions. If they are broad, we should expect large Allocation zones. In this case, the *Graham* non-obviousness standard can usually be trusted to link expected reward to relative R&D costs for most – and in extreme cases all – inventions regardless of cleverness. In this world, even minimal doctrines may be enough to satisfy our third policy goal. Conversely, narrow distributions imply large Default zones. In this case, the

Graham standard leads to 50:50 allocations even when one invention is markedly cleverer than the other. Here, legislators and judges will often decide that it is better to create additional Allocation zones that reduce or eliminate the number of Default outcomes. This typically means adopting affirmative defenses that implement additional PHOSITA-X tests for very clever inventions.¹³²

Which world do we actually live in? In principle, the verdict distribution can be measured. For example, scholars could play the same videotaped patent trial to 100 average Americans and ask them to deliver verdicts. This experiment, of course, has never been done. Still, we can guess at the answer. Given that we are all prospective jurors, this might involve little more than asking ourselves how consistently we would rank, say, a PHOSITA-75 invention above its PHOSITA-50 counterpart. We can also find hints in the empirical literature. Naively, we expect parties to avoid the expense of litigating defenses in the Default Zone. After all, why argue an issue that you are likely to lose?¹³³ By that measure, it worth noting that just under one-third of all patent cases tried in the 1990s litigated the issue of non-obviousness to judgment.¹³⁴ This suggests a Goldilocks result in which verdict distributions have significant but not overwhelming breadth.

In what follows, we will assume that our verdict distributions are narrow enough to create substantial Default Zones. This provides a strong argument for creating defenses that invite juries to conduct additional comparative cleverness determinations far from PHOSITA-50.

E. Implications for Doctrine

Our Literal Infringement example is clearly a special case. Despite this, it already contains important lessons for doctrine. First, it reminds us of unintended consequences. Recall that the Supreme Court invented the *Graham* standard to address our first policy goal, *i.e.*, screen out inventions not worth the “embarrassment of an exclusive patent.” In the process, however, it inadvertently created a process for allocating reward among successive inventors. Though facially reasonable, this rule falls short of our benchmark Absolute Scope rule. We will see that

¹³² We return to this subject in Section 8.

¹³³ See note 122 above. Practicing litigators will immediately point out that this argument is simplistic. For example, lawyers may deliberately run up opponents’ costs by forcing them to prove obvious facts.

¹³⁴ Sean M. McElowney, “New Insights on the ‘Death’ of Obviousness: An Empirical Study of District Court Obviousness Opinions,” 2006 *Stan. Tech. L. Rev.* 4 (2006) at Table 2; *But see*, Benjamin H. Graf, “Prognosis Indeterminable: How Patent Non-Obviousness Outcomes Depend Too Much on Decisionmakers,” 9 *Cardozo Pub. L. Pol’y & Ethics J.* 567, 595 (2011) (PTO examiner interviews show that non-obviousness determinations “are largely a function of the decision-maker”).

this defect can be ameliorated – though not eliminated – by grafting additional PHOSITA-X inquiries onto doctrine.

Finally, we have assumed so far that litigation is a single, one-shot event. But many patents face multiple litigations. Under principles of collateral estoppel this means that a patent which fails to establish *Graham* non-obviousness in, say, the third litigation will lose *every* lawsuit thereafter. More generally, patent owners facing multiple litigations will receive systematically less expected income than they would if their claims were tried in a single action. It is hard to see how this problem can be solved short of abolishing collateral estoppel altogether.

7. How Doctrine Allocates Reward (B): Literal Infringement and Blocking Patents

Blocking patents do a surprisingly good job of allocating reward among successive inventors. If our third policy goal were everything, we would expect the same 50:50 rule to operate throughout patent law. However, we must also consider our second goal. It insists that the 1G patent monopoly should have limited duration and breadth. This implies that sufficiently clever 2G inventors should sometimes be allowed to build on 1G insights and still escape infringement.

This Section extends our probabilistic patents analysis to include the Doctrine of Equivalents. Part A sets the stage by reviewing the sometimes incoherent case law that surrounds the Doctrine of Equivalents. We argue that attempts to rationalize the Doctrine will almost certainly require juries to assess the 2G invention’s cleverness. Part B analyzes how a rationalized Doctrine would allocate reward for different 2G cleverness standards. Significantly, we find that very strong versions of the Doctrine of Equivalents approximate our Blocking Patents rule. Part C concludes by discussing the prospects for combining Literal Infringement and the Doctrine of Equivalents within a single unified doctrine.

A. The Doctrine of Equivalents.

The Doctrine of Equivalents was originally designed to prevent copyists from using trivial changes to evade verbal “metes and bounds” descriptions of the patent monopoly.¹³⁵ This eventually led to a “Function-Way-Result” Test which asked whether the accused device “performs substantially the same function in substantially the same way to achieve

¹³⁵ Kenneth D. Bassinger “Unsettled Expectations in Patent Law: *Festo* and the Moving Target of Claim Equivalence,” 48 *How. L.J.* 685, 691 (2005) (“Recognizing that language is often a dull instrument with which to define the intellectual nuances of an invention, the courts created the doctrine of equivalents to provide an escape from the confines of the restrictive literalism of claim language”).

substantially the same result.” But how are judges to define “substantially”? If the answer is “more than *de minimis*,” any changes beyond trivial, calculated evasion should escape infringement. Precedents that describe the Doctrine as a deterrent against “pirating,” “mere imitation,” “insubstantiality,”¹³⁶ “minor variations,” and “unimportant and insubstantial changes”¹³⁷ tend to reinforce this view. More recently, however, the US Supreme Court has emphasized that the doctrine is “not limited” to cases of copying and piracy but instead defines “the notion of identity between a patented invention and its equivalent.”¹³⁸ This seems to recognize expansive (if vague) judicial statements that 2G inventors cannot escape so long as “the gist” of their invention is the same and there are no “unexpected or substantially different results.”¹³⁹ These standards bring the Doctrine much closer to innovation theory’s familiar concept of patent breadth.

That said, the content of words like “equivalent” and “gist” remains elusive. Historically, most efforts to clarify the Doctrine have involved arguments that modifications that use an “interchangeable” element cannot be substantial and are therefore “equivalent.”¹⁴⁰ “Interchangeability,” in turn, is almost always framed in terms of what “a skilled artisan” would expect.¹⁴¹ This implies a cleverness test and bears an obvious resemblance to *Graham’s* PHOSITA standard. At the same time, the “skilled artisan” phraseology pre-dates and remains distinct from the usual formula of a “Person Having Ordinary Skill in the Art.” Maddeningly, the nature of these differences is obscure. On the one hand, assertions that the notional artisan

¹³⁶ *Id.* at 691, 695.

¹³⁷ Alan L. Durham, “Patent Symmetry,” 87 *B.U.L. Rev.* 969, 1004-05 (2007).

¹³⁸ *Warner Jenkinson Co., Inc. v. Hilton Davis Chemical Co.*, 520 U.S. 17, 34-35 (1997).

¹³⁹ *Atlas Powder*, 750 F.2d at 1580 and note 3.

¹⁴⁰ Frustratingly, the test is not absolute. *Chiuminatta Concrete Concepts, Inc. v. Cardinal Indus., Inc.*, 145 F.3d 1303, 1309 (Fed. Cir. 1998) (“Moreover, a finding of known interchangeability, while an important factor in determining equivalence, is certainly not dispositive.”); see also Paul R. Michel, “The Role and Responsibility of Patent Attorneys in Improving the Doctrine of Equivalents,” 40 *Idea* 123 (2000) (remarking that “known interchangeability” is currently “a mere factor” instead of “a true test”).

¹⁴¹ Cases that invoke the concept of a skilled artisan without using the traditional “Person Having Ordinary Skill in the Art” standard include *Graver Tank*, 339 U.S. at 609 (“An important factor ... is whether a person reasonably skilled in the art would have known of the interchangeability of an ingredient not contained in the patent with one that was”); *Interactive Pictures Corp. v. Infinite Pictures, Inc.*, 274 F.3d 1371, 1382-83 (Fed. Cir. 2002) (“The known interchangeability test looks to the knowledge of a skilled artisan to see whether that artisan would contemplate the interchange as a design choice.”). Commentators routinely paraphrase these cases as invoking a PHOSITA standard. See, e.g. Bassinger, *supra* at 695. Durham notes that the US Supreme Court has variously invoked “persons skilled in the art” and “skilled practitioners” and also asked whether whether the 2G substitution was “routine” or else required “a further advancement in the art.” Durham, *supra*, at 1011-12.

must have “reasonable skill” hint that Doctrine’s cleverness threshold could be lower than the PHOSITA’s “average” or “ordinary” test.¹⁴² On the other, statements that the “advancement” must be more than “routine”¹⁴³ point in the opposite direction.

Suggestions for reforming the Doctrine usually fall into three categories.¹⁴⁴ The first would leave the Function-Way-Result Test unchanged but demand more explicit argument and evidence to prove it.¹⁴⁵ This would presumably weaken the Doctrine while avoiding any policy discussion of what patent breadth is or ought to be.¹⁴⁶ The second proposal would expand “known interchangeability” into a “true test.”¹⁴⁷ While the proposal’s details are obscure, it is reasonable to think that the word “known” would continue to be measured by the awareness of skilled artisans. This clearly implies a PHOSITA-X threshold test, though not necessarily a PHOSITA-50 one. The final and arguably most popular suggestion would be to identify breadth with *Graham*’s non-obviousness standard. This approach is already widely used overseas¹⁴⁸ and has been endorsed by various US commentators¹⁴⁹ and judges.¹⁵⁰ As the late Judge Nies argued in a celebrated concurrence, “A substitution in a patented invention cannot be both

¹⁴² Durham, *supra*, 87 *B.U.L. Rev.* at 1011.

¹⁴³ *Id.* at 1011.

¹⁴⁴ Paul R. Michel, “The Role and Responsibility of Patent Attorneys in Improving the Doctrine of Equivalents,” 40 *IDEA* 123 (2000); *see also*, M. Scott Boone, “Defining and Refining the Doctrine of Equivalents: Notice and Prior Art, Language, and Fraud,” 43 *IDEA* 645, 649-654 (2003).

¹⁴⁵ *Id.* According to Judge Michel, the approach would build on *Malta v. Schulmerich Carillons, Inc.* 952 F.2d 1320, 1327 (1991) (criticizing “offhand and conclusory statements” that left jury “to its own imagination on the technical issue of equivalency”) and *Lear-Siegler, Inc. v. Sealy Mattress Co.*, 873 F.2d 1422 (1989) (jury must receive “particularized testimony and linking argument” for each element of Triple Identity Test).

¹⁴⁶ More explicit evidence and argument by counsel might, of course, lead to clearer judicial decisions.

¹⁴⁷ *See, e.g., Michel, supra; Boone, supra*, at 654.

¹⁴⁸ Mark D. Janis, “Who’s Afraid of Functional Claims? Reforming the Patent Law’s § 112, ¶ 6 Jurisprudence,” 15 *Santa Clara Computer & High Tech. L.J.* 231, 267 (1999).

¹⁴⁹ *See, e.g., Durham, supra* 87 *B.U.L. Rev.* 969; Scott P. Zimmerman, “The Doctrine of Equivalents: A Call for Congressional Reinvigoration,” 40 *IDEA* 599 (2000); T. Whitley Chandler, “Prosecution History Estoppel, the Doctrine of Equivalents, and the Scope of Patents,” 13 *Harv. J. Law & Tec.* 465, 489-95 (2000); Michael T. Siekman, “The Expanded Hypothetical Claim Test: A Better Test for Infringement for Biotechnology Patents Under the Doctrine of Equivalents,” 2 *B.U. J. Sci. & Tech. L.* 6 (1996).

¹⁵⁰ Michel, *supra*. *See also* Helen Nies’ opinion in *Roton Varrier, Inc. v. Stanley Works*, 79 F.3d 1112, 1128 (Fed Cir. 1996).

nonobvious and insubstantial.”¹⁵¹ Significantly, Judge Nies may not have realized that *Graham’s* PHOSITA-50 standard is just one choice among many. In fact, her argument applies to any PHOSITA-X value greater than PHOSITA-50.¹⁵²

B. Analysis

Given these uncertain precedents, it is hard to predict how judges will ultimately reform and clarify the Doctrine. At the same time, practically all courts invoke PHOSITA-like cleverness judgments. No reform is likely to change this. The only question is whether the corresponding PHOSITA-X threshold will be large or small. This section examines the various possibilities and how they change the baseline allocations generated by *Graham’s* non-obviousness test.

We begin by reminding the reader of our Literal Infringement/Blocking Patents analysis. In that case, doctrine asked juries whether the 1G and 2G patents were non-obvious. This meant that both patents were compared against the same PHOSITA-50 standard and, implicitly, each other. The Doctrine of Equivalents adds a second standard: In addition to *Graham* non-obviousness, jurors must also compare the 2G invention’s cleverness against the 1G patent’s breadth. This second, breadth threshold can take on any PHOSITA-X value from 1 to 100. For reasons that appear below, it is enough to consider two examples.

Narrow Breadth. Fig. 3 shows how much reward the 2G invention receives where breadth is defined by a PHOSITA-25 standard. Unlike previous Figures, we plot the 2G invention’s cleverness against its expected allocation share. For clarity, the Figure depicts the particular case in which the 1G invention’s verdict distribution peaks at PHOSITA-50.

We start on the far left-hand side. For very low cleverness inventions the 2G patent has almost no chance of exceeding the PHOSITA-25 threshold. For this reason, it earns nothing in expectation. This leads to a 100:0 Default Zone.¹⁵³ Now examine the far right hand side. Here,

¹⁵¹ *Roton, supra*, 79 F.3d at 1128.

¹⁵² See note 155 and accompanying text, *infra*.

¹⁵³ The label is a matter of convenience. Since 1G patent has a 50 percent chance of failing the non-obviousness test, 0:0 outcomes are equally likely.

inventions that are significantly cleverer than PHOSITA-25 escape the 1G patent and pay

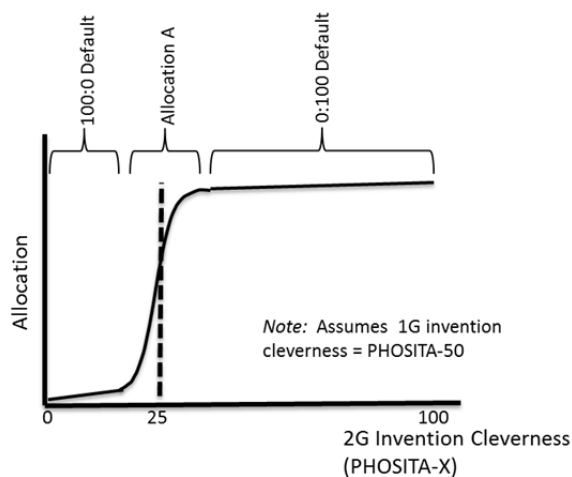


Fig. 3

nothing. This implements our second policy goal but also leads to very extreme 0:100 allocations that almost certainly fail to reflect the parties' relative R&D costs. We label this the 0:100 Default Zone in Fig. 3.

Finally, consider the middle region near PHOSITA-25. Unlike our blocking patents case, the 1G patent is measured against *Graham*'s PHOSITA-50 standard while the 2G patent is measured against our PHOSITA-25 breadth definition. This distorts the calculation so that the 2G inventor receives roughly three times as much for her cleverness as the 1G inventor.¹⁵⁴ This allocation can only be right for rare cases where R&D costs rise steeply at each successive development step.

So far, we have assumed that the 1G invention's verdict distribution peaks at PHOSITA-50. If the the 1G patent is significantly cleverer than the *Graham* non-obviousness juries will almost certainly decide that it is valid. In this case, the 2G invention's expected allocation share will be smaller at point in Fig. 3. Similarly, less clever 1G invention will produce higher 2G rewards.

High PHOSITA-X. We now consider the opposite case in which most follow-on inventions infringe the 1G patent. Fig. 4 depicts this situation for the PHOSITA-75 case. As in Fig. 3, we depict the specific case where the 1G invention's verdict distribution peaks at PHOSITA-50.

¹⁵⁴ To see this, consider the case where the 1G invention's verdict distribution peaks at PHOSITA-50 and the 2G curve peaks at PHOSITA-25. Then we expect the 1G patent to be upheld 50% of the time. Even in these cases, however, there is still a 50% chance that the 2G invention will exceed PHOSITA-25 and escape infringement. We therefore expect the final allocation to be 25:75 on average.

Not surprisingly, everything to the left of the PHOSITA-75 standard resembles our Blocking Patent analysis in Section 6. On the left hand side there is a 100:0 Default Zone where the 2G invention is far below the PHOSITA-50 threshold for blocking patents. This leads to a region

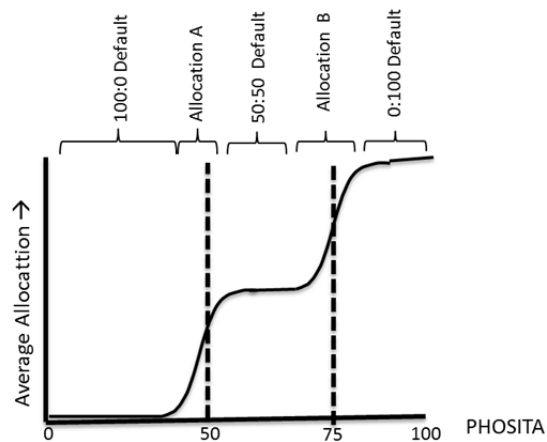


Fig. 4

near the *Graham* threshold (Allocation Zone A) where revenue is allocated in proportion to relative cleverness. There is also a second, 50:50 Default Zone for 2G inventions whose verdict distributions fall in the broad middle region between PHOSITA-50 and PHOSITA-75.¹⁵⁵

Conversely, the area at and to the right of our PHOSITA-75 standard resembles our Low Breadth case. This includes a probabilistic region (“Allocation A”) near the 2G invention and a 0:100 region for very clever 2G inventions. As before, making the 1G invention more clever would reduce the 2G invention’s expected allocation at each point in the figure.

Putting these observations together, we see that everything to the left of the PHOSITA-75 threshold replicates our “Blocking Patents” allocation rule. Indeed, the two doctrines are identical for infinite (PHOSITA-100) breadth.

¹⁵⁵ The zone disappears in the special case where breadth and non-obviousness standards are identical. As Prof. Adams remarks, combining the two standards sets up a syllogism. If the 1G patent would enable a PHOSITA to make the 2G invention, the latter falls within the 1G patent and should not be patentable. But if the 1G patent does not enable a PHOSITA to make the 2G invention, the latter is patentable and falls outside the 1G patent’s breadth. Charles W. Adams, “Allocating Patent Rights Between Earlier and Later Inventions,” 54 *S. Louis L.J.* 55, 60 (2009). This may explain why most blocking patent cases involve Literal Infringement and not the Doctrine of Equivalents.

C. Lessons for Doctrine

The deep divide between Literal Infringement and The Doctrine of Equivalents outcomes is puzzling. Certainly, nothing in innovation economics suggests that inventors who add new claims elements require fewer incentives than those who rethink existing ones. But this leaves us with an uncomfortable choice. Either the two rules share some underlying unity, or one must surely wrong.

The fact that large PHOSITA-X versions of the Doctrine of Equivalents produce much the same allocation outcomes as Literal Infringement provides an important clue. If the two doctrines really do converge, the difference may be less about substance than judicial economy. After all, cleverness inquiries require expensive jury determinations. But how clever can a 2G invention be if it fails to escape the 1G patent's literal claims language? Judges could well decide that it is better to outlaw cleverness inquiries in Literal Infringement cases altogether. Like all bright line rules, this is bound to produce mistakes. Indeed, Section 8A will discuss one such example at length. Still, this may not matter much if defenses like the "Reverse Doctrine of Equivalents" provide a safety valve for radically clever improvements.¹⁵⁶

The doctrinal fissure is harder to rationalize for low PHOSITA-X versions of the Doctrine of Equivalents. Here, Literal Infringement really does produce markedly different outcomes. In theory, Congress could re-unify doctrine by abolishing Literal Infringement and mandating a single Doctrine of Equivalents test for all cases. But in that case what PHOSITA-X value should Congress use to define breadth? In an ideal world, legislators could estimate the optimal PHOSITA-X thresholds for our second and our third goals separately and then split the difference. In reality, this procedure would consist mostly of guesswork.

8. How Doctrine Allocates Reward (C): Defenses

We saw in Section 6 that any patent doctrine based on the *Graham*'s PHOSITA test automatically establishes baseline rules for allocating reward among inventors. But we also argued that these results could be improved by adopting doctrinal architectures that invite juries to make additional inquiries. This Section reviews various infringement defenses that have been proposed over the years and compares their potential for improving *Graham*'s baseline allocation rule.

¹⁵⁶ See Section 8A, below.

We begin by reviewing the often controversial defenses that courts have developed to address multi-inventor issues. These include the Reverse Doctrine of Equivalents (Part A), the Pioneer Patents Doctrine (Part B) and the Written Description Requirement (Part C). We then compare these doctrines against our third goal.

A. Reverse Doctrine of Equivalents

The Reverse Doctrine of Equivalents was introduced more than 100 years ago¹⁵⁷ and remains (apparently¹⁵⁸) good law today. However, it is expressly limited to literal infringement cases¹⁵⁹ and is seldom applied.¹⁶⁰ Despite this, it remains surprisingly popular among scholars.¹⁶¹ This suggests that it could be readily revived given a clear policy reason to do so.

¹⁵⁷ *Westinghouse v. Boyden Power Brake Co.*, 170 US 537, 568 (1898).

¹⁵⁸ The *Westinghouse* rule languished for a over half a century before the Supreme Court revived it in *Graver Tank & Mfg. Co. v. Linde Air Prods., Inc.*, 339 U.S. 605 (1950). Thereafter, it remained uncontroversial, if seldom used, for another fifty years. This changed in 2002 when the Federal Circuit suggested that the *Graver Tank* holding had been superseded by 35 USC §112. If so, the reverse doctrine could no longer be invoked to “constrict[] the literal language when it is clearly claimed.” *Tate Access Floors, Inc. v. Interface Architectural Res., Inc.*, 279 F.3d 1357, 1368 (Fed. Cir. 2002). However, this argument has been widely criticized. Mark A. Lemley, “The Changing Meaning of Patent Terms,” 104 *Mich. L. Rev.* 101 (2005) (arguing that Reverse Doctrine of Equivalents is distinct from written description because it is measured at the time of infringement rather than filing); Cyrill P. Rigamonti, “Conflicting Theories of Equivalence: 35 USC §112, Para 6 in the Supreme Court and the Federal Circuit” 40 *IDEA* 163 (2000); *but see* Martin J. Adelman and Gary L Francione, “The Doctrine of Equivalents in Patent Law: Questions that *Pennwalt* did not Answer,” 137 *U. Pa. L. Rev.* 673, 724. (1989)(“Section 112 “should be viewed as a legislative codification of the reverse doctrine of equivalents.”). Moreover, more recent Federal Circuit decisions continue to mention the defense. *See, e.g., Biogen, Inc. v. Berlex Labs, Inc.* 318 F.3d 1132 (Fed. Cir. 2003); *Plant Genetic Sys., NV v. DeKalb Genetics Corp.*, 315 F.3d 1335 (Fed. Cir. 2003); *Amgen, Inc. v. Hoechst Marion Roussel, Inc.* 314 F.3d 1313 (Fed. Cir. 2003); *DePuy Spine, Inc. v Medtonic Sofamor Daenk, Inc.* 567 F.3d 1314 (Fed Cir. 2009) (“The Supreme Court has recognized it to be a viable defense, even if it is rarely asserted”); *Wavetronix v. EIS Elec. Integrated Systems*, 573 F.3d 1343, 1360-61 (Fed Cir. 2009) (noting that accused system worked in a much different way).

¹⁵⁹ *Texas Instruments, Inc., v. United States International Trade Commission*, 846 F.2d 1369, 1371 (Fed Cir. 1988). We will see that the Pioneer Patents doctrine provides similar rules in Doctrine of Equivalents cases.

¹⁶⁰ *But see, Gardner v. Ford Motor Co.* 1990 U.S. Dist. LEXIS 21052 (WD Wash. 1999) (second invention was far removed from the “principle, structure, and operation” of claimed invention); *U.S. Steel Corp. v. Phillips Petroleum*, 865 F.2d 1247, 1252, n. 9 (Fed Cir. 1989); *American Standard, Inc. v. Pfizer, Inc.* 722 F. Supp. 86, 104 (D. Del. 1989) (endorsing “principle”-based test and criticizing tests based on whether invention functions “in a substantially different way”); *Smithkline Diagnostics, Inc. v. Helena Labs. Corp.*, 859 F.2d 878, 889 (Fed. Cir. 1988) (finding that the defendant could establish non-infringement by showing that his device performs a similar function in a substantially different way); *Precision Metal Fabricators Inc. v. Jetstream Systems Co.*, 693 F. Supp. 814, 819 (N.D. Cal. 1988) (infringement did not occur where defendant’s machines did “not operate on the same principle” and “[w]hatever similarities exist are incidental and do not enhance the operation of defendants' machines.”); *Leesona Corp. v. United States*, 530 F.2d 896, 906 (Ct. Claims 1976) (It is well settled that more than a literal response to the terms of the claims must be shown to make out a case of infringement.”)

Strangely, most of what we know about the Doctrine still comes from the original *Westinghouse* opinion. The case involves – and describes in prolix, obsessive detail – air brakes for trains. Westinghouse had developed a system in which each car’s brakes were driven by a local tank which was, in turn, pressurized and controlled by lines from the locomotive. This required an elaborate system of “triple” and “auxiliary valves.” Boyden invented a new triple valve that made the auxiliary valves unnecessary. Despite remarkable cleverness, however, the new arrangement still fell within the literal language of Westinghouse’s patent. This might have encouraged the Court to declare a blocking patent. In fact, the Court went much further by declaring that Boyden’s invention had “so far changed the principle of the device that the claims of the patent, literally construed, have ceased to represent his actual invention.”¹⁶² Left unsaid was how much change would be required to trigger the defense. Modern case law¹⁶³ and scholarship¹⁶⁴ says that the change must be “radical.” However, the matter is not free from

¹⁶¹ Extended discussions and proposed extensions of the doctrine can be found in, *inter alia*, R. Scott Roe, Nanotechnology: When Making Something Smaller is Nonobvious,” 12 *B.U. J. Sci. & Tech. L.* 127 (2006); Michael A. Carrier, “Cabining Intellectual Property Through a Property Paradigm,” 54 *Duke L.J.* 1 (2004); Christopher D. Hazuka, “Supporting the Work of Lesser Geniuses: An Argument for Removing Obstructions to Human Embryonic Stem Cell Research,” 57 *U. Miami L. Rev.* 157 (2002); Anand Gupta, “Patent Law: The Supreme Court Reinforces the Validity of the Doctrine of Equivalents in *Warner-Jenkinson Co. v. Hilton Davis Chemical Co.*,” 23 *S. Ill. U. L. J.* 123 (1998); Mark A. Lemley “The Economics of Improvement in Intellectual Property Law,” 75 *Tex. L. Rev.* 989 (1997); Robert Merges, Intellectual Property Rights and Bargaining Breakdown: The Case of Blocking Patents, 62 *Tenn. L. Rev.* 75, 101-03 (1994); Michael S. Greenfield, “Recombinant DNA Technology: A Science Struggling with the Patent Law,” 44 *Stan. L. Rev.* 1051 (1992); Laura A. Handley, Refining the Graver Tank Analysis With Hypothetical Claims: A Biotechnology Exemplar,” 5 *Harv. J. Law & Tech.* 31(1991).

¹⁶² *Westinghouse*, *supra* 170 US 537, 568. The Supreme Court famously reaffirmed the rule in 1951: “Thus, where a device is so far changed in principle from a patented article that it performs the same or a similar function in a substantially different way, but nevertheless falls within the literal words of the claim, the doctrine of equivalents may be used to restrict the claim and defeat the patentee’s action for infringement.” *Graver Tank & Mfg. Co., Inc. v. Linde Air Prods. Co.*, 339 U.S. 605, 608 (1950).

¹⁶³ *Scripps Clinic & Rsch. Foundation v. Genentech, Inc.* 927 F.2d 1565, 1581 (Fed. Cir. 1991) (suggesting that a literally infringing device may escape liability if it represents a “radical improvement on the patented technology”).

¹⁶⁴ Peter Lee, “The Accession Insight and Patent Infringement Remedies,” 110 *Mich. L. Rev.* 175, 185-186 (2011) (radical improvements); R. Scott Roe, Nanotechnology: When Making Something Smaller is Nonobvious, 12 *BU J. Sci & Tech. L.* 127, 141 (2006) (doctrine prevents 1G patent owners from stifling “radical improvements.”); Simone A. Rose, “On Purple Pills, Stem Cells, and Other Market Failures: A Case for a Limited Compulsory Licensing Scheme for Patent Property,” 48 *How. L.J.* 579, 605 (2005) (doctrine is available for “infringing acts...which produce radically pioneering inventions”); Christopher D. Hazuka, “Supporting the Work of Lesser Geniuses: An Argument for Removing Obstructions to Human Embryonic Stem Cell Research 57 *U. Miami L. Rev.* 157, 212 (2002); Janice M. Mueller, “No ‘Dilettante Affair’: Rethinking the Experimental Use Exception to Patent Infringement for Biomedical Research Tools,” 76 *Wash. L. Rev.* 1 (2001) at 45 and n. 228 (Reverse Doctrine of Equivalents triggered by “radical improvement”); Maureen O’ Rourke, “Toward a Doctrine of Fair Use in Patent Law,” 100 *Colum. L. Rev.* 1177, 1193 (2000) (Reverse Doctrine of Equivalents operates where 2G inventor is “radical improver” of underlying invention); Robert P. Merges, “A Brief Note on Blocking Patents and the Reverse Doctrine of Equivalents in Biotechnology Cases,” 73 *J. Pat & Trademark Off. Soc’y* 878, 883 (1991) (safety valve that prevents patent owners from stifling radical improvements).

doubt and some scholars have suggested that the 2G advance need only be radical compared to the underlying 1G invention.¹⁶⁵ Still others argue that mere “substantial” or “significant” changes suffice.¹⁶⁶ In terms of our present analysis, then, existing precedent would support any 2G cleverness test ranging from “substantial” to “radical.”¹⁶⁷

The Court also suggested an alternative theory. The outcome might have been different, the justices argued, if Westinghouse’s method “would naturally have suggested the device” invented by Boyden.¹⁶⁸ But in fact, “the Westinghouse patent, *if [Boyden] had had it before him*, would scarcely have suggested the method he adopted to accomplish those results. Under such

¹⁶⁵ Alan Devlin, “The Misunderstood Function of Disclosure in Patent Law,” 23 *Harv. J. Law & Tec* 401, 434 and note 166 (2010) (Reverse Doctrine of Equivalents operates when subsequent invention represents “an overwhelming technological leap” beyond the original patent disclosure); Carolyn Abbot and David Booton, “Using Patent Law’s Teaching Function to Introduce an Environmental Ethic Into the Process of Technical Innovation,” 21 *Geo. Int’l Envtl. L. Rev.* 219 (2009) at 249 and note 130 (“exceeds the contribution made by the patented invention”); Mark A. Lemley, The Economics of Improvement in Intellectual Property Law, 75 *Tex. L. Rev.* 989, 1012 (1997) (doctrine applies to “sufficiently radical” improvers and is “most likely” where the “value of the improvement exceeds the value of the original invention”).

¹⁶⁶ Donald S. Chisum, “*Bilski v. Kappos*: Everything Old is New Again: Weeds and Seeds in the Supreme Court’s Business Method Patents Decision: New Directions for Regulating Patent Scope,” 15 *Lewis & Clark L. Rev.* 11, 24 (2011) (“The prevailing standard for equivalency – substantial change – can take into account whether, on the one hand, an accused equivalent represents a merely inconsequential design around, or on the other, presents a significant invention”); *see also*, Ben Depoorter, “The Several Lives of Mickey Mouse: The Expanding Boundaries of Intellectual Property Law” 9 *Va. J.L. & Tech.* 4, 51 (2004) (Reverse Doctrine of Equivalents requires “a significant contribution that takes the invention outside of the original, allegedly infringed-upon patent”); Alan Devlin, “Patent Law’s Parsimony Principle,” 25 *Berkeley Tech. L.J.* 1693, 1738 and note 220 (2011) (“sufficiently large leap forward”); Andrew Wasson, Protecting the Next Small Thing: Nanotechnology and the Reverse Doctrine of Equivalents, 2004 *Duke L. & Tech. Rev.* 10 (changes must be “sufficiently different”); *see also Texas Instruments, Inc. v. U.S. Int’l Trade Comm’n*, *supra*, 846 F.2d at 1371 (doctrine applies where defendant’s device is “sufficiently different” from plaintiff’s).

¹⁶⁷ Although most commentators stress cleverness, a few argue that value should be measured by economic success or performance. These user-centered concepts hint at a utility analysis. Christopher D. Hazuka, “Supporting the Work of Lesser Geniuses: An Argument for Removing Obstructions to Human Embryonic Stem Cell Research,” 57 *U. Miami L. Rev.* 157, 212 (arguing that Reverse Doctrine of Equivalents should be measured by “scientific” and “economic” advances) (2002); Mark A. Lemley, “The Economics of Improvement in Intellectual Property Law,” 75 *Tex. L. Rev.* 989, 1065 (1997) (arguing that Reverse Doctrine of Equivalents should apply where improvement “is such a major advance ...that its value is much greater than the original”). Alan Devlin, “Patent Law’s Parsimony Principle,” 25 *Berkeley Tech. L.J.* 1693, 1739 (2011) (arguing that “significant leaps” can be measured by whether the underlying patent’s technology is “commercially defunct.”) Doctrinally, this would transform utility from the current threshold requirement to a full-blown UTILITY-X comparison. US law has long disfavored this notion. N. Scott Pierce, “*In re Dane K. Fisher*: An Exercise in Utility” 6 *J. High Tech. L.* 1 (2006) (survey of case law and scholarship distinguishing between “positive utility” from “degree”).

¹⁶⁸ *Westinghouse*, *supra* 170 U.S. at 573.

circumstances, the law entitles him to the rights of an independent inventor.”¹⁶⁹ This suggests that a 2G innovator should be allowed to claim a defense where she “can truly assert that her claimed invention was not made possible (*i.e.* enabled) by what was described by the prior inventor.”¹⁷⁰ This sounds very much like a rule that 2G inventors should escape infringement where the 1G patent possessed no option value for their specific invention.

B. Pioneer Patents

The Reverse Doctrine of Equivalents grew out of an earlier rule that “pioneer” inventions should receive broader patent protection than mere “improvements.”¹⁷¹ Unlike the Reverse Doctrine of Equivalents, the Pioneer Patents concept expressly extends to non-literal infringement under the Doctrine of Equivalents. Like the Reverse Doctrine, its status is unclear. While the Federal Circuit seemed to overrule it in the late 1980s,¹⁷² more recent courts¹⁷³ and commentators¹⁷⁴ continue to recognize it.

There is also a split of authority on how the rule should be phrased. Most courts hold that pioneer patents must represent a radical improvement over prior art. This standard is variously expressed as requiring a “broad breakthrough,” “major advance,” “basic operational concept,” “broadly new,” “devoid of significant prior art,” “primary,” “basic,” “generic,” “original,” or “key.”¹⁷⁵ However, some courts insist that the “pioneer” and “improvement” concepts only

¹⁶⁹ *Id.* at 573 (emphasis supplied).

¹⁷⁰ Robert A. Bohrer and John T. Prince, “A Tale of Two Proteins: The FDA’s Uncertain Interpretation of the Orphan Drug Act,” 12 *Harv. J. Law & Tech.* 365, 407 (1999). The view is also implicit in arguments that the Reverse Doctrine of Equivalents is an equitable doctrine designed to prevent unjust windfalls. Laura A. Handley, “Refining the Graver Tank Analysis With Hypothetical Claims: A Biotechnology Exemplar,” 5 *Harv. J. Law & Tech.* 31, 49-50 (1991).

¹⁷¹ *Westinghouse* at 561-562.

¹⁷² *Tex. Instruments, Inc. v. United States ITC*, 846 F.2d 1369 (Fed Cir. 1988) (purporting to overrule doctrine); see generally, Brian J. Love, “Interring the Pioneer Invention Doctrine,” 90 *N.C.L. Rev.* 379, 395-404 (2012).

¹⁷³ See Brian J. Love, “Interring the Pioneer Invention Doctrine,” *supra* at 395-404 and cases cited therein.

¹⁷⁴ Favorable discussions of the doctrine include John R. Thomas, “The Question Concerning Patent Law and Pioneer Inventions,” 10 *Berkeley High Tech. L.J.* 35 (1995); Michael J. Meurer and Craig Allen Nard, “Invention, Refinement and Patent Claim Scope: A New Perspective on the Doctrine of Equivalents,” 93 *Geo. L.J.* 1947, 1989 (2005), and Georgia E. Kralovic, “The Principle of Fair Notice: Is It Prudent Guidance for the Future of Patent Law?,” 26 *Pepp. L. Rev.* 89, 104-05 (1998) (remarking that pioneer status is “given little weight,” but “make[s] sense intuitively.”)

¹⁷⁵ Thomas, *supra* at 48 and cases cited therein.

mark the ends of a continuum¹⁷⁶ so that breadth should be continuously adjusted for even small differences in cleverness. These alternatives are similar to those at issue in Reverse Doctrine cases and should presumably be decided the same way.

C. Written Description

In its most general form, the Written Description requirement dates from the Supreme Court's 1822 pronouncement that a patent applicant has an obligation "to describe what his own improvement is, and to limit [the] patent to such improvement."¹⁷⁷ However, this function became redundant with the rise of claims pleading and lay "dormant for many years."¹⁷⁸ In 1967, the Court of Customs and Patent Appeals' *Ruschig* decision¹⁷⁹ held that a patent whose disclosure potentially explained how to make "something like half a million possible compounds"¹⁸⁰ could not be amended to add a specific drug called chloropromamide. The reason, the court explained, was that the inventor had provided no "motivation for wanting to make the compound in preference to others" and failed to disclose the compound "as something appellants actually invented."¹⁸¹

On the face of things, the *Ruschig* rule was limited to the narrow question of when amended claims could relate back to the original filing date.¹⁸² In 1997, however, the Federal Circuit's *Eli Lilly* decision broadened the Written Description requirement to include original claims. In particular, it held that inventors trying to patent a particular gene had to disclose an atom-by-atom description of the corresponding DNA molecule.¹⁸³ Most commentators interpreted this

¹⁷⁶ See *Hughes Aircraft Co. v. United States*, 717 F.2d 1351, 1362 (Fed. Cir. 1983); *Cont'l Oil Co. v. Cole*, 634 F.2d 188, 198 n.7 (5th Cir. 1981); *Price v. Lake Sales Supply R.M., Inc.*, 510 F.2d 388, 394 (10th Cir. 1974). *Shields v. Halliburton Co.*, 667 F.2d 1232, 1238 (5th Cir. 1982); *Swanson v. Unarco Indus., Inc.*, 479 F.2d 664, 669 (10th Cir. 1973); *Corning Glass Works v. Anchor Hocking Glass Corp.*, 374 F.2d 473, 476 (3d Cir. 1967); *Kori Corp. v. Wilco Marsh Buggies & Draglines, Inc.*, 561 F. Supp. 512, 522 (E.D. La. 1981); *Ronson Corp. v. Maruman of Cal., Inc.*, 224 F. Supp. 479, 481 (S.D. Cal. 1963).

¹⁷⁷ *Evans v. Eaton*, 20 US (7 Wheat.) 356, 433-34 (1822)

¹⁷⁸ Dan L. Burk and Mark A. Lemley, "Policy Levers in Patent Law," 89 *Va. L. Rev.* 1575, 1653 (2003).

¹⁷⁹ *Application of Ruschig*, 379 F.2d 990 (1967).

¹⁸⁰ *Id.* at 1555.

¹⁸¹ *Id.* at 1558.

¹⁸² *Moba BV v. Diamond Automation, Inc.*, 325 F.3d 1306, 1319 (2003) (doctrine is designed to "prevent the addition of new matter to claims.")

¹⁸³ *Regents of the University of California v. Eli Lilly & Co.*, 119 F.3d 1559, 1566 (Fed Cir. 1997).

as a new and *ad hoc* “super-enablement” requirement for biotechnology inventions.¹⁸⁴ Formally, however, the Court seemed to be announcing the more general principle that courts should ignore literal infringement when the 1G patent failed to show that the inventor had actually “possessed” the 2G invention. This view has since been confirmed by cases that apply the doctrine to sofa design and other non-biotech technologies.¹⁸⁵

Lilly was widely criticized by scholars¹⁸⁶ and some Federal Circuit judges.¹⁸⁷ In 2010, the Federal Circuit wrote the *en banc Ariad* decision to rationalize and explain the defense.¹⁸⁸ The case turned on a drug patent that purported to claim all compounds that interfered with the binding of certain molecules. The Court held that the Written Description requirement barred this patenting of unknown drugs. Conceding that the “possession” test had “never been very enlightening,” the Court added that the patent had to “describe an invention understandable to [a] skilled artisan” and, further, “show that the inventor invented the invention claimed.” This would prevent patent owners from “merely recit[ing] a description of the problem to be solved while claiming all solutions to it and ... cover any compound later actually invented and determined to fall within the claim’s functional boundaries, leaving it to the pharmaceutical industry to complete an unfinished invention.”¹⁸⁹

Barring Supreme Court intervention, *Ariad* will likely remain definitive for decades. For our purposes, it makes three important points. First, Written Description is a universal defense that applies to literal and non-literal infringement alike. Second, Written Description focuses on what the 1G inventor actually knew. This is subtly different from the Reverse Doctrine of Equivalents, which asks whether the 2G inventor would have gained new insights by reading the 1G patent. Finally, the test is objective and asks how a PHOSITA would have interpreted the 1G inventor’s disclosure. This presumably guards against obscure hints that could only be understood by a genius-level (*e.g.* PHOSITA-95) inventor.

¹⁸⁴ Christopher M. Holman, “Is Lilly Written Description a Paper Tiger? A Comprehensive Assessment of the Impact of Eli Lilly and its Progeny in the Courts and PTO,” 17 *Alb. L.J. Sci. & Tech.* 1 at 19 and note 91 (2007).

¹⁸⁵ *The Gentry Gallery, Inc. v. The Berkline Corp.*, 134 F.3d 1473 (Fed. Cir. 1998) (sofa armrest controls).

¹⁸⁶ Christopher M. Holman, *supra*, 17 *Alb. L.J. Sci. & Tech.* 1 at 17-18.

¹⁸⁷ *Id.* at 18.

¹⁸⁸ *Ariad Pharmaceuticals, Inc. v. Eli Lilly & Co.*, 598 F.3d 1336 (Fed. Cir. 2010) (*en banc*).

¹⁸⁹ *Id.* at 47-48.

D. Analysis (Pt. 1): Option Value Revisited

The Reverse Doctrine of Equivalents, Pioneer Patent, and Written Description defenses ask juries to make a second PHOSITA-X determination beyond *Graham*'s baseline PHOSITA-50 inquiry. We now ask how well these additional inquiries narrow the gap between doctrine and our third social goal.

We begin by asking which defenses best fit our option value theory of utility. We have already said that option value often depends 2G inventors' ability to spot unusually clever (e.g. PHOSITA-95) applications and improvements. The Reverse Doctrine captures such cases by asking whether the 1G patent would have helped the 2G inventor achieve her insight or, conversely, "would scarcely have suggested the method ... adopted." By comparison, the Written Description test is narrowly focused on what the 1G inventor actually knew and communicated in terms that would have been understandable to a PHOSITA-50. This necessarily excludes hints to 2G geniuses that often possess significant option value. Judges who accept our option value arguments should want to replace Written Description with a reinvigorated Reverse Doctrine of Equivalents defense.

D. Analysis (Pt. 2): Allocation Rules

We have seen that courts and commentators have interpreted the Reverse Doctrine of Equivalents, Pioneer Patents, and Written Description defenses in different ways over the years. Here, we examine the various interpretations and ask which one best serves our third social goal.

Absolute Cleverness. Some courts have argued that the Reverse Doctrine of Equivalents and Pioneer Patent defenses should apply whenever the 2G patent is exceptionally clever in absolute terms. However, this ignores situations where the 1G patent's cleverness (and, implicitly cost) is comparable to or exceeds the 2G patent's. Courts that accept our analysis should reject these decisions out of hand.

Direct Comparison. Some versions of the Reverse Doctrine of Equivalents and Pioneer Patent defenses hold that the 2G patent should escape infringement when its cleverness is only slightly larger than the 1G patent's. Readers should think of this as a case-specific threshold set to coincide with the peak of the 1G invention's verdict distribution. After that, the analysis is the same as our discussion of strong Doctrine of Equivalents rules.¹⁹⁰ In particular, we expect the defense to produce improved allocations whenever the 1G and 2G inventions are comparably

¹⁹⁰ See Fig. 4 and accompanying text, *supra*.

clever. Conversely, inventions that possess markedly different cleverness will usually receive 50:50 (or 0:100) allocations without regard to cleverness or relative cost.

Radical Improvements. Most Reverse Doctrine of Equivalents and Pioneer Patent cases specify that the 2G invention must represent a “radical” improvement over its 1G predecessor to

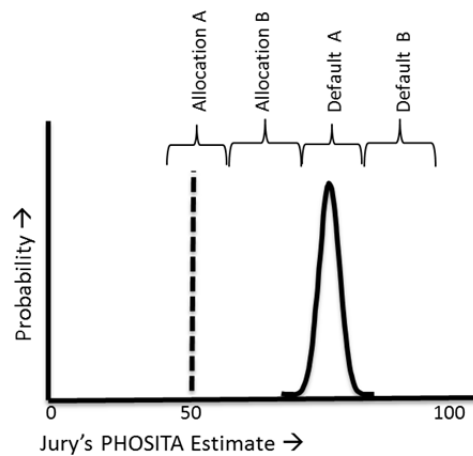


Fig. 5

escape infringement. Fig. 5 summarizes the situation. Reading from right to left, the PHOSITA-50 standard establishes a probabilistic zone (“Allocation A”) that ranges from 50:50 to 0:100 as before. Now, however, there is a second probabilistic zone (“Allocation B”) where the 2G patent is radically cleverer than its 1G predecessor. Because the radical improvement rule does *not* apply where the two inventions are similarly clever, our first Default A zone is centered on the 2G patent. This puts the 50:50 default blocking outcome exactly where it should be: Where the two inventions are already known to have similar cleverness. Finally, there may sometimes be a second (“Default B”) zone that provides 0:100 allocations in cases where the 2G patent is much cleverer than its 1G predecessor.¹⁹¹

Based on the foregoing, we argue that Pioneer Patent/Reverse Doctrine of Equivalents defenses that require “radical” improvement rules are the best way to implement our third policy goal. Courts that accept our analysis should discard the Written Description test and concentrate on reviving these doctrines.

¹⁹¹ We should not insist on such radical improvements that the new Allocation B zone overlaps the original PHOSITA-50 region and becomes redundant. This could happen, for example, if juries interpreted “radical” improvement in such extreme terms that the defense was only available to, say, PHOSITA-99 inventions.

9. Doctrine vs. Theory: Assessing the Shortfall

We have now completed our survey of how patent doctrine implements modern neoclassical theory. On the positive side, we have seen that doctrine does a reasonable job of tracking our three neoclassical goals. At the same time, any patent system that includes *Graham*'s PHOSITA test is unavoidably imperfect. This leaves us in an awkward position. If the current system worked exceptionally well – or badly – we would immediately know whether to keep it. Instead, the question stands on a knife edge.

This Section begins by recapitulating the strengths and weaknesses of existing doctrine. We then ask whether a radically different system that discarded the PHOSITA proxy and tried to estimate R&D costs directly could do a better job.

A. How Good is the Proxy?

We have argued that the PHOSITA standard is best understood as a proxy for R&D cost. Like all proxies, this introduces various possible errors:

Imperfect Proxy. No proxy is perfect. We have seen that there are a significant number of cases where cleverness says little or nothing about R&D cost. In other cases, incremental changes in cleverness can receive disproportionately large (or small) rewards.

Ignoring Profit. Because our PHOSITA proxy only tracks relative R&D costs, it cannot detect the moment when each inventor recoups her absolute cost and begins to earn a profit. For this reason, no PHOSITA-based doctrine can ever fully implement our Absolute Scope rule by redistributing profit upstream.

Systematic Error. Our probabilistic PHOSITA mechanism often generates 50:50 or 0:100 default allocations that are divorced from cleverness or, implicitly, cost. Adding a Reverse Doctrine of Equivalents defense mitigates but does not solve this problem.

Under-Rewarding Multi-Step Inventions. Technologies based on multiple patents often lead to multiple lawsuits. In these cases, collateral estoppel will truncate patent rewards compared to a single, one-shot lawsuit.

A Cushion for Mistakes? We have argued that any *Graham*-based doctrine is inherently imperfect. That said, policymakers can still influence which kinds of mistakes it makes. More specifically, it is always better to have an over-priced invention than no invention at all. This suggests that legislators and judges should err on the side of broad patent monopolies and

liberal rewards. The argument is even stronger in the multi-inventor case, where breaking just one link in a long development chain can kill promising technologies.

The good news is that current doctrine allocates reward on the basis of *relative* costs. This means that each inventor will still be able to cover her costs provided that the *absolute* reward is large enough. At the same time, a deliberately inflated reward violates our second policy goal. Given current state of the art, any attempt to adjust this tradeoff is largely a matter of guesswork.

Industry-Specific Rules. This article has pursued the traditional intuition that global rules are attractive and worthwhile. But real judges do not need to be purists. Instead, industry-specific rules can and do provide a safety valve when global doctrine fails to fit to conditions on the ground.

Pharmaceutical R&D provides a spectacular example of such ad hoc-ery.¹⁹² This is not surprising since the industry often features (a) inventions that have no immediate use, and (b) R&D costs that typically rise several-fold for each step along the development pipeline. These unusual facts make industry-specific rules particularly attractive. At the same time, Big Pharma is uniquely well-funded and able to protect its interests. How would we know if other, poorer industries were encountering similar problems? Unlike industry-specific rules, a sensible global rule protects rich and poor alike. For this reason, judges should adopt *ad hoc* solutions reluctantly and then only as a last resort.

C. The Road Not Taken

It is not enough to say that the PHOSITA standard is “good” or “bad.” We must also ask “Compared to what?” Here, the obvious alternative is asking courts to estimate R&D cost directly. Doctrinally, this means moving the allocation inquiry from liability to damages. Existing case law suggests at least two ways to do this.

Apportionment. The most direct solution is to apportion revenue according to each patent’s respective “contribution” to products purchased by end-users. The idea is not as strange as it sounds. Indeed, some district courts apportioned damages in the Nineteenth and early Twentieth Centuries¹⁹³ while Congress debated the idea as recently as 2009.¹⁹⁴ There is also

¹⁹² For example, courts may be more willing to apply the Reverse Doctrine of Equivalents in biology than other industries. See, Laura A. Handley, Refining the Graver Tank Analysis with Hypothetical Claims: A Biotechnology Exemplar,” 5 *Harv. J. Law & Tec.* 31, 45 - 50 (1991).

¹⁹³ Amy L. Landers, “Patent Claim Apportionment, Patentee Injury, and Sequential Innovation,” 19 *Geo. Mason L. Rev.* 471, 494-504 (2012).

considerable scholarly support for the idea. Prof. Landers makes a compelling argument that “reasonable royalty” damages would divide revenues more fairly among patent owners.¹⁹⁵ Similarly, Prof. Lee argues that the related equitable doctrine of “accession” creates “a middle zone of flexibility” between 0:100 and 100:0 outcomes. This might “embolden” courts to allocate reward differently between pioneer and improving inventors.¹⁹⁶

From our standpoint, neither of these proposals is quite right. The reason is that they focus on each patent’s “contribution” in providing value (*i.e.* utility) to end-users. This concept is decidedly different from asking what costs each inventor incurred to advance the technology. Even so, shifting the allocation inquiry to R&D costs sounds like a minor detail.

Compulsory Licenses. Compulsory licenses provide a second way to apportion revenue. Prof. Tur-Sinai has argued that this is the best way to implement the Absolute Scope rule.¹⁹⁷ However, modern courts are notoriously hostile to the idea.¹⁹⁸ That said, the concept remains fairly mainstream and continues to appear in academic articles.¹⁹⁹

New Headaches. The foregoing approaches would require courts to estimate corporate R&D costs directly. Naively, this sounds straightforward. On reflection, though, one could say the same thing about public utility commissions’ efforts to set fair rates of return. In practice, these inquiries almost always require large specialized bureaucracies and make frequent errors. As the US Supreme Court has pointed out in the antitrust context, regulation is a “daunting task”²⁰⁰ which courts should avoid.²⁰¹

¹⁹⁴ Megan L. Wiggins, “Patent Reform and Damages Apportionment: Addressing the Concerns of Industry-Scale Users of the U.S. Patent System Without Legislatively Mandating a ‘Specific Contribution over the Prior Art,’” 40 *Seton Hall L. Rev.* 273, 276-277 (2010).

¹⁹⁵ Amy L. Landers, *supra*, at 504 - 509.

¹⁹⁶ Peter Lee, *supra*, 110 *Mich. L. Rev.* 175 at 237, 239.

¹⁹⁷ Ofer Tur-Sinai, *supra*, 50 *IDEA* 723 (2010).

¹⁹⁸ Merges and Nelson, *supra* at 840.

¹⁹⁹ For a recent survey of the idea in both theory and practice, see Carol M. Nielsen and Michael R. Samdzija, “Compulsory Patent Licensing: Is It a Viable Solution in the United States?” 13 *Mich. Telecomm. Tech. L. Rev.* 509, 533-36 (2007); see also, Simone A. Rose, *supra*, 48 *How. L.J.* 579, 618-24 (advocating “limited” compulsory licenses where “market failure” exists).

²⁰⁰ *Verizon Comms. Inc. v. Law Offices of Curtis V. Trinko*, 540 US 398, 414 (2004)

²⁰¹ *Id.* at 415.

Replacing Graham. We have argued that any doctrine that includes *Graham*'s non-obviousness test automatically contains a baseline allocation rule for successive inventors. On the other hand, our first social goal still requires a threshold standard: If we overrule *Graham* we must replace it with something. The least radical solution is to retain *Graham* and only estimate costs for whatever patents survive the non-obviousness screen. However, this would only improve incentives for 2G inventions that are much cleverer than PHOSITA-50. A ambitious reform would replace *Graham* itself. Under this rule, R&D projects that incurred above-average costs would be deemed non-obvious. This, however, would expand the inquiry still further: In addition to estimating the parties' own R&D costs, the court would also have to estimate the average cost of similar projects throughout the industry.

10. Conclusion

Nearly a half century has passed since the Supreme Court built neoclassical economics into the foundations of patent doctrine. But while innovation economics has moved on, most legal scholars still treat the literature as a grab bag of conflicting theories. This gloomy assessment seems to put global rules out of reach. It is also unreasonable. Far from being fragmented, the various neoclassical insights share the same fundamental assumptions and methods. For this reason, they are best seen as special cases of a deeper and more general theory. This article has identified and explored three social goals that any such theory must have. Two of these are generalized versions of principles (non-obviousness, reward size) that were already familiar in the 1960s. The third specifies how patent reward should be divided among successive inventors.

Remarkably, current doctrine uses *Graham*'s PHOSITA standard to address all three goals. This practically guarantees that attempts to address one goal will have unintended impacts on the others. We have argued that separate PHOSITA-X standards can relax this link for our first and second principles. Future attempts to reform the Doctrine of Equivalents should take advantage of this fact. At the same time, *any* doctrine based on a PHOSITA non-obviousness test *automatically* implements economically inefficient allocations. A revived Reverse Doctrine/Pioneer Patents defense would go some distance toward narrowing this gap.

No doctrine that includes the PHOSITA concept can ever be perfect. Full implementation of our third social goal would require judges to scrap the current system in favor of estimating corporate R&D costs directly. This is bound to be a large and error-prone undertaking. By comparison, the PHOSITA standard – despite its flaws – is at least reasonably accurate and can be implemented by judges and juries. Radical change, if it comes at all, should be left to Congress.