

Florida State University Law Review

Volume 39 | Issue 3

Article 4

2012

Better Mistakes in Patent Law

Andres Sawicki
0@0.com

Follow this and additional works at: <http://ir.law.fsu.edu/lr>



Part of the [Law Commons](#)

Recommended Citation

Andres Sawicki, *Better Mistakes in Patent Law*, 39 Fla. St. U. L. Rev. (2012).
<http://ir.law.fsu.edu/lr/vol39/iss3/4>

This Article is brought to you for free and open access by Scholarship Repository. It has been accepted for inclusion in Florida State University Law Review by an authorized administrator of Scholarship Repository. For more information, please contact bkaplan@law.fsu.edu.

FLORIDA STATE UNIVERSITY LAW REVIEW



BETTER MISTAKES IN PATENT LAW

Andres Sawicki

VOLUME 39

SPRING 2012

NUMBER 3

Recommended citation: Andres Sawicki, *Better Mistakes in Patent Law*, 39 FLA. ST. U. L. REV. 735 (2012).

BETTER MISTAKES IN PATENT LAW

ANDRES SAWICKI*

ABSTRACT

This Article analyzes patent mistakes—that is, mistakes made by the patent system when it decides whether a particular invention has met the patentability requirements. These mistakes are inevitable. Given resource constraints, some might even be desirable. This Article evaluates the relative costs of patent mistakes, so that we can make better ones.

Three characteristics drive the costs of mistakes: their type (false positive or false negative), timing (early or late), and doctrinal basis (utility, novelty, nonobviousness, and so on). These characteristics make some mistakes more troubling than others.

This Article compares the costs of making mistakes of different types, at different times, and on different doctrinal bases. These comparisons produce some surprising results—for example, under certain plausible conditions, it will be better to wrongly refuse to grant a patent than to wrongly invalidate a patent that had already been granted. The conclusions here have important implications for persistent issues in patent law, including how closely courts should scrutinize the validity of issued patents and how the Patent and Trademark Office should allocate scarce enforcement resources.

I. INTRODUCTION.....	736
II. PATENT THEORY, PATENT DOCTRINE, AND PATENT MISTAKES.....	740
A. <i>The Goals of Patent Law</i>	741
B. <i>Patentability Rules</i>	742
C. <i>Mapping the Terrain of Patent Mistakes</i>	744
1. <i>Type</i>	746
2. <i>Timing</i>	746
3. <i>Doctrine</i>	749
III. HOW TO MAKE BETTER MISTAKES.....	750
A. <i>False Positives</i>	750
1. <i>Doctrine as a Direct Selection Mechanism</i>	751
2. <i>Timing Effects</i>	753
(a) <i>Disclosure Rules</i>	753
(b) <i>Definiteness</i>	754
(c) <i>Scope Rules</i>	755
(d) <i>Invention Rules</i>	757
3. <i>Summary and Caveats</i>	758
B. <i>False Negatives</i>	760
1. <i>Doctrine as an Indirect Selection Mechanism</i>	761
(a) <i>Value-Based Variation</i>	763

* Bigelow Teaching Fellow and Lecturer in Law, The University of Chicago Law School. For very helpful comments and conversations, the author would like to thank Robert Ahdieh, Douglas Baird, Anya Bernstein, Alexander Boni-Saenz, Anthony Casey, Lee Fennell, Bernard Harcourt, William Hubbard, Aziz Huq, Mark Janis, Saul Levmore, Doug Lichtman, Jonathan Masur, Anthony Niblett, Randal Picker, Eric Posner, Naomi Schoenbaum, Julia Simon-Kerr, Lior Strahilevitz, and Matthew Tokson, as well as workshop participants at Emory University School of Law, Florida International University School of Law, Indiana University Maurer School of Law, Ohio State University Moritz College of Law, University of Arizona Rogers College of Law, University of Chicago Law School, University of Maryland Carey School of Law, University of Miami School of Law, University of Notre Dame Law School, Wake Forest University School of Law, and the 2011 IP Scholars Conference. My deepest thanks go to my wife, Jessica, for her unwavering support and encouragement throughout. Any mistakes (better or worse) are, of course, my own.

(b) <i>Industry-Specific Variation</i>	765
2. <i>Timing Effects</i>	767
(a) <i>Observability of False Negatives</i>	767
(b) <i>Availability of Non-Patent Appropriation Mechanisms</i>	768
(i) <i>Trade Secrecy</i>	769
(ii) <i>Tacit Knowledge</i>	771
(iii) <i>Trademarks</i>	773
(iv) <i>Contracts</i>	775
3. <i>Summary and Caveats</i>	776
IV. IMPLICATIONS.....	777
A. <i>The Relative Costs of Patent Mistakes</i>	777
B. <i>Resource Allocation</i>	778
C. <i>The Presumption of Validity</i>	779
V. CONCLUSION.....	780

I. INTRODUCTION

The patent system makes many mistakes, frequently granting patents that should be denied and denying patents that should be granted.¹ One approach is to introduce reforms aimed at reducing the rate of these mistakes.² But even if these reforms are successful, mistakes are ultimately inevitable. A rate of zero mistakes is both unrealistic and undesirable.³

But not all mistakes are created equal. This Article thus asks: how should we get patent law wrong? In other words, given that the patent system will make some mistakes, which ones should we prefer?

The patent system includes two government institutions that evaluate patentability: the Patent and Trademark Office and the federal courts. The focus here is on the patent system's mistaken appli-

1. For examples of the rich literature emphasizing the desirability of getting things right more often, see JAMES BESSEN & MICHAEL J. MEURER, *PATENT FAILURE: HOW JUDGES, BUREAUCRATS, AND LAWYERS PUT INNOVATORS AT RISK* (2008); ADAM B. JAFFE & JOSH LERNER, *INNOVATION AND ITS DISCONTENTS: HOW OUR BROKEN PATENT SYSTEM IS ENDANGERING INNOVATION AND PROGRESS AND WHAT TO DO ABOUT IT* (2004); James Gleick, *Patently Absurd*, N.Y. TIMES MAG., Mar. 12, 2000, at 44, 47, available at <http://www.nytimes.com/2000/03/12/magazine/patently-absurd.html?pagewanted=print&src=pm>; Michael A. Heller & Rebecca S. Eisenberg, *Can Patents Deter Innovation? The Anticommons in Biomedical Research*, 280 SCIENCE 698 (1998); Mark Lemley, Douglas Lichtman & Bhaven Sampat, *What to Do About Bad Patents?*, 28 REGULATION 10 (2005); Douglas Lichtman & Mark A. Lemley, *Rethinking Patent Law's Presumption of Validity*, 60 STAN. L. REV. 45 (2007); Robert P. Merges, *As Many as Six Impossible Patents Before Breakfast: Property Rights for Business Concepts and Patent System Reform*, 14 BERKELEY TECH. L.J. 577 (1999); and John R. Thomas, *Collusion and Collective Action in the Patent System: A Proposal for Patent Bounties*, 2001 U. ILL. L. REV. 305 (2001).

2. See, e.g., Beth Simone Noveck, *"Peer to Patent": Collective Intelligence, Open Review, and Patent Reform*, 20 HARV. J.L. & TECH. 123 (2006) (arguing for mechanisms to allow public participation in patent examination so as to reduce the number of wrongly-issued patents).

3. See Mark A. Lemley, Essay, *Rational Ignorance at the Patent Office*, 95 NW. U. L. REV. 1495 (2001) (arguing that the costs of early mistakes are too low on average to justify the expense of avoiding all early mistakes).

cations of the patentability rules. These rules are a set of doctrines that the patent system uses to decide whether an inventor is entitled to a patent. Three characteristics might affect the costs of these mistakes. First, the mistakes might be false positives (incorrect grants of patents) or false negatives (incorrect denials of patents). Second, they might be made early, during the evaluation of the patent application, or late, during a challenge to an issued patent's validity in infringement litigation. And third, they might be made with respect to any of the patentability rules: subject matter, utility, novelty, nonobviousness, enablement, best mode, or definiteness.⁴

The goal of this Article is to provide a relative sense of how the costs of mistakes vary along with these three characteristics.⁵ A conclusive assessment depends on the resolution of difficult (perhaps intractable) empirical questions, but we can make some progress with the theoretical perspective taken here. The analysis leads to four general conclusions.

First, false positives on different doctrines will necessarily cause problems of different kinds. The costs of failing to detect violations of some rules will therefore be greater than the costs of failing to detect violations of others. The patent system's failure to detect violations of the nonobviousness requirement, for example, will create patent thickets—areas in which there are many overlapping rights to an invention. On the other hand, the patent system's failure to detect violations of the enablement requirement will force the public to waste resources duplicating the inventor's achievement. These two kinds of problems—overlapping rights and wasteful duplication—will naturally impose costs of different magnitudes; they are, after all, different kinds of problems. False positives on some doctrines will therefore be more costly than false positives on others.

Second, the timing of a false positive will affect its costs on a doctrine-by-doctrine basis. Late false positives necessarily follow early ones, so the analysis depends on the additional costs incurred when the patent system produces a late false positive after an early one. For some doctrines—like enablement and definiteness—late false positives will add little to the costs incurred because of the early false positive. For such doctrines, we should be (roughly) indifferent to late

4. I set aside written description for reasons set forth *infra* note 25.

5. This Article's approach is similar to that taken in Michael J. Meurer, *Patent Examination Priorities*, 51 WM. & MARY L. REV. 675 (2009). Meurer, however, considers only mistakes made during the PTO's evaluation of patent applications. *Id.* This Article also includes mistakes made during litigation and thus adds a critical dimension to the analysis: the timing of the mistake. In a related vein, Mark Lemley has argued that the costs of an average early false positive are likely to be low, and thus it is not worthwhile to spend additional resources reducing the rate of early mistakes. See Lemley, *supra* note 3. I depart from that analysis by assessing variations in costs along the three characteristics of type, timing, and doctrinal basis.

false positives.⁶ But for other doctrines—like subject matter and non-obviousness—late false positives may impose significant additional costs. The patent system should therefore approach these doctrines differently, being more cautious of late false positives on subject matter and nonobviousness and less worried about late false positives on enablement and definiteness.

False negatives present somewhat different considerations. The third general conclusion is that, unlike for false positives, the doctrinal basis of a false negative mistake will only indirectly affect the cost of the mistake, if it all. The cost of a false negative mistake flows from the disincentive effect it will have on future innovators who worry that they too will be wrongly denied patent rights. This disincentive effect is not necessarily doctrine-specific; there is nothing in the doctrines themselves that dictates, for example, that the disincentive effect from wrongly denying a patent for failure to comply with the enablement requirement is larger than the disincentive effect from wrongly denying a patent for failure to comply with the novelty requirement. As a result, the doctrinal basis of the false negative may only indirectly affect the costs of the mistake.

It is possible that particular patentability rules tend to misfire on patents with importantly different characteristics. One possibility is that the doctrines vary in the rate at which they produce false negatives on patents covering inventions of different values. If so, then false negatives on doctrines that err on patents covering high-value inventions will be costlier than false negatives on doctrines that err on patents covering low-value inventions. Even if there is no variation on this dimension, the doctrines might differ on other grounds. Most plausibly, the doctrines will vary in the rate at which they produce false negatives in different industries; to the extent that they do, false negatives will have industry-specific effects by doctrine.

Finally, the comparison between early and late false negatives depends on the relative importance of three factors: (1) the efficacy of a short period of exclusivity; (2) the efficacy of non-patent appropriation mechanisms (and their sensitivity to the timing of their adoption); and (3) the visibility of false negatives. Late false negatives afford the inventor at least a short period of exclusivity (from patent grant to invalidation) in which to appropriate some of the value of her investment; when that period is lucrative, late false negatives will be less costly to the inventor than early ones. But the two other factors will favor early false negatives. Inventors will turn to non-patent appropriation mechanisms to solve the public good problem created by false negatives. These mechanisms include trade secrecy,

6. Late false positives are of course undesirable for all the reasons we wish the patentability rules to be properly enforced as a general matter.

tacit knowledge, trademarks, and contracts. To varying degrees, these mechanisms are more easily implemented early in the product development lifecycle than later. So early false negatives—those made during the patent system’s evaluation of the application—should generally be easier for inventors to overcome than late false negatives—those made during the patent system’s evaluation of an issued patent during infringement litigation. Moreover, because it will be easier for third parties to learn of late false negatives than early ones, late mistakes may have larger negative effects on incentives to innovate than early ones.

This is a highly uncertain and complex area of the law. The story here cannot be conclusive; these are ultimately empirical questions that can only be settled with empirical data. But given the well-known difficulties in obtaining good empirical data on the effects of the patent system,⁷ it is better to proceed on the basis of well-developed intuitions. The intuitions developed here can provide new ways of thinking about patent doctrines, inform patent policy, and guide future empirical research.

This Article proceeds as follows. Part I provides background on the goals of patent law and describes how the patentability rules further those goals. It then introduces a framework for assessing the costs of patent mistakes. Part II uses that framework to reach four general conclusions about the costs of mistakes of different types, made at different times, and made with respect to different patentability rules. Part III draws out the implications of this analysis. Part IV concludes.

7. See, e.g., John F. Duffy, *Rules and Standards on the Forefront of Patentability*, 51 WM. & MARY L. REV. 609, 618 (2009) (noting in the context of debates about the scope of patentable subject matter that “the ultimate policy judgment—the extent to which the potentially positive effects of patents are outweighed by their potential negative effects—has long been recognized as unknown given the current state of human knowledge”). In Fritz Machlup’s 1958 formulation:

If we did not have a patent system, it would be irresponsible, on the basis of our present knowledge of its economic consequences, to recommend instituting one. But since we have had a patent system for a long time, it would be irresponsible, on the basis of our present knowledge, to recommend abolishing it.

FRITZ MACHLUP, SUBCOMM. ON PATENTS, TRADEMARKS, AND COPYRIGHTS OF THE S. COMM. ON THE JUDICIARY, 85TH CONG., AN ECONOMIC REVIEW OF THE PATENT SYSTEM 80 (Comm. Print 1958). Though we’ve made progress in the last fifty years, there is still a great deal of uncertainty. See Jonathan M. Barnett, *Property as Process: How Innovation Markets Select Innovation Regimes*, 119 YALE L.J. 384, 386 (2009) (“[I]t is probably uncontroversial among most economically informed observers that Machlup’s qualified statement still characterizes our current understanding of the net social value of the intellectual property system as a general matter.”).

II. PATENT THEORY, PATENT DOCTRINE, AND PATENT MISTAKES

This Part provides the building blocks for analyzing patent mistakes, which I define as the mistakes that the patent system makes when it applies the doctrines governing patentability to particular inventions.⁸ These doctrines, which I call the patentability rules, are designed to determine whether granting a patent would further the patent system's goals. I therefore begin by exploring those goals in Section A. I then describe the patentability rules and their relationships to the patent system's goals in Section B. With that foundation in place, Section C identifies the three primary characteristics that affect the costs of patent mistakes.

Patent commentators debate many aspects of what patent law should do and how patent law does whatever it is that it should do. In order to simplify the exposition, I discuss only the areas of broadest agreement. The details might change depending on how we resolve disputed questions of patent theory. The overall picture, however, will be similar regardless of how those debates play out. I take here as given that patent law is primarily designed to provide incentives to innovate and secondarily designed to encourage disclosure of technical and legal information.⁹ Though those points might be dis-

8. My focus here is on what may be termed retail mistakes; that is, those that occur on a case-by-case basis. I set aside wholesale mistakes that occur for groups of patents as a whole. An example of a wholesale mistake would be an interpretation of the subject matter requirement that wrongly excluded an entire technological field, although note that the subject matter doctrine can also be applied in a retail manner. See *Bilski v. Kappos*, 130 S. Ct. 3218, 3225-29 (2010) (rejecting "two proposed categorical limitations" for deciding whether an invention covers patentable subject matter and urging courts to take a case-by-case approach instead). Wholesale mistakes may raise different concerns than the retail mistakes I evaluate here. Cf. Samuel L. Bray, *Preventive Adjudication*, 77 U. CHI. L. REV. 1275, 1314-33 (2010) (developing a distinction between retail sorting, in which the government decides on a case-by-case basis, and wholesale sorting, in which the government decides for a large group of cases). Of course, the patent system surely also makes mistakes in deciding, for example, questions of infringement. These are outside the scope of this Article.

9. This, at least, is the standard account. See, e.g., WILLIAM M. LANDES & RICHARD A. POSNER, *THE ECONOMIC STRUCTURE OF INTELLECTUAL PROPERTY LAW* 294 (2003) ("The standard rationale of patent law is that it is an efficient method of enabling the benefits of research and development to be internalized, thus promoting innovation and technological progress."); Dan L. Burk, *The Role of Patent Law in Knowledge Codification*, 23 BERKELEY TECH. L.J. 1009, 1010 (2008) ("[T]he dominant justification for the patent system has shifted toward an economic rationale based upon incentives."); Dan L. Burk & Mark A. Lemley, *Policy Levers in Patent Law*, 89 VA. L. REV. 1575, 1580 (2003) ("There is virtually unanimous agreement that the purpose of the patent system is to promote innovation by granting exclusive rights to encourage invention."); Jeanne C. Fromer, *Patent Disclosure*, 94 IOWA L. REV. 539, 548 (2009) (describing patent law as "rewarding inventors for taking two steps . . . : to invent in the first instance and to reveal information to the public about these inventions"); Gideon Parchomovsky & R. Polk Wagner, *Patent Portfolios*, 154 U. PA. L. REV. 1, 12 (2005) ("The standard justification . . . is that patents are necessary to solve an appropriability problem . . . stem[ming] from the 'public good' characteristics of intellectual goods."). Here are the traditional citations to the Constitution's Intellectual Property Clause and to the Supreme Court's description of the "economic philosophy"

puted, they are the foundation for a general consensus about patent law. To the extent that we reach other conclusions, the discussion here would have to be modified accordingly.

A. *The Goals of Patent Law*

Patent law aims to provide optimal incentives to innovate.¹⁰ Without patents, inventors deciding whether to develop an invention face an appropriability problem. Research produces valuable information that has public good characteristics—it is both non-excludable (that is, the inventor cannot easily prevent strangers from using it) and non-rival (that is, one person's use of the information does not limit another person's use of it).¹¹ Once the inventor completes her research, her rivals will be able to copy the invention and sell it at much lower prices because they have not incurred the inventor's cost of development. The inventor will therefore find it difficult to profit from her research projects.

The patent system tries to solve this appropriability problem by granting inventors a right to exclude others from their inventions.¹² That right to exclude encourages inventors to conduct costly research by promising the ability to charge supracompetitive prices for their inventions. To the extent that patents in fact confer monopolies, though, we will incur their social costs. The patent system's solution thus pays the costs of monopolies to gain the benefits of increased incentives to invent.

If it were costless to do so, the patent system could evaluate each patent application to directly determine whether the trade-off was worth it from a social welfare perspective.¹³ We could ask something like, "All things considered, will society be better off if we grant this patent or deny it?" But that evaluation would be prohibitively costly. The patent system instead uses a set of rules that aim—at some

underlying it. See U.S. CONST. art. I, § 8, cl. 8; *Mazer v. Stein*, 347 U.S. 201, 219 (1954). There are, of course, other views about what the patent system does or should do, but I set them aside and focus on the standard rationale here. For some sample alternative accounts, see LANDES & POSNER, *supra*, at 326-32 (arguing that patent law serves in large part to channel innovators away from trade secret law so that information about inventions reaches the public domain more quickly); Clarisa Long, *Patent Signals*, 69 U. CHI. L. REV. 625 (2002) (arguing that firms use patents as signals of their knowledge capital); and Parchomovsky & Wagner, *supra* (arguing that patents are used to form portfolios that function primarily through scale and diversity effects even though individual patents might have negative expected value).

10. See *supra* note 9.

11. Parchomovsky & Wagner, *supra* note 9, at 12-13.

12. Burk & Lemley, *supra* note 9, at 1580 (noting the standard view that "exclusive rights address the public goods nature of inventions that are expensive to produce but easy to appropriate").

13. Edmund W. Kitch, *The Nature and Function of the Patent System*, 20 J.L. & ECON. 265, 281-82 (1977).

reasonable cost—to sort applications that should be granted from those that should be denied, given the background goals of patent law. This sorting is the “central problem” of the patent system.¹⁴ The next Section describes the patentability rules used to solve that sorting problem.

B. Patentability Rules

There are many plausible interpretations of what, precisely, the patentability rules should do and whether they achieve their intended purposes in a reasonable manner. Again, I set aside these debates about the means and ends of the patentability rules and stick instead to the standard accounts.¹⁵ Though the particulars are open to question, the general outlines here are less controversial.

The patentability rules can be placed into four categories. First, there are rules regarding scope: the subject matter and utility doctrines.¹⁶ The point of these rules is to withhold patents when inventors seek them so early in the innovation process that they would permit control over too broad a range of follow-on innovation. Thus, the subject matter doctrine permits patents on “anything under the sun that is made by man,” but it gets its teeth by excluding “laws of nature, physical phenomena, and abstract ideas.”¹⁷ Patents on the latter would grant excessive control over downstream inventions.¹⁸ Because some inventions that pass the subject matter test may still excessively inhibit subsequent research, the utility doctrine denies patents on inventions for which the only known use is as a subject of scientific inquiry.¹⁹

Second, there are rules regarding the invention itself: novelty and nonobviousness.²⁰ These invention rules are designed to avoid issuing

14. *Id.* at 280.

15. For an example of debates about the means and ends of the patentability rules, compare Mark D. Janis, *On Courts Herding Cats: Contending with the “Written Description” Requirement (and Other Unruly Patent Disclosure Doctrines)*, 2 WASH. U. J.L. & POL’Y 55, 62-69 (2000) (arguing “that no compelling reason to recognize a distinct written description requirement exists”), with Michael Risch, *A Brief Defense of the Written Description Requirement*, 119 YALE. L.J. ONLINE 127 (2010) (arguing that a distinct written description requirement ensures that the patent’s scope is calibrated to match the inventor’s achievement).

16. 35 U.S.C. § 101 (2006).

17. *Diamond v. Chakrabarty*, 447 U.S. 303, 309 (1980); *accord Bilski v. Kappos*, 130 S. Ct. 3218, 3225 (2010).

18. *O’Reilly v. Morse*, 56 U.S. 62, 113 (1853) (rejecting a claim as covering unpatentable subject matter because it would allow the inventor to “shut[] the door against inventions of other persons”); *Burk & Lemley*, *supra* note 9, at 1643.

19. *Brenner v. Manson*, 383 U.S. 519, 534-35 (1966); *Burk & Lemley*, *supra* note 9, at 1644-46 (explaining that the *Brenner* rule is driven by concerns that “giving patent protection too early—before the actual use of the product has been identified— . . . might deter research by others on the use of the product”).

20. 35 U.S.C. §§ 102-103 (2006).

patents on things that we do not need the patent system in order to get. Novelty requires that the patent claim something that is not already known.²¹ Without something like the novelty requirement, society would pay the price of patents without any corresponding benefits in return—after all, the public already knew about the claimed invention, so there is no longer any need to provide an incentive for someone to invent it.²² Nonobviousness goes one step further: it holds that even if the public did not already know about the precise invention claimed in the patent, it is still ineligible for patent protection if it would have been obvious to a person in the field.²³ The idea here is that inventors will make obvious improvements to existing technology because these improvements are available “off the shelf” and require no supracompetitive returns as an inducement for inventors to attain them. Even if rivals can copy the invention, it is so easy (and therefore cheap) for the inventor to reach it that the public good problem would not pose a serious obstacle to this kind of progress.²⁴

Third, there are rules regarding disclosure: the enablement and best mode doctrines.²⁵ The point of the disclosure rules is to force the

21. 35 U.S.C. §§ 101-102; Sean B. Seymore, *Rethinking Novelty in Patent Law*, 60 DUKE L.J. 919, 922 (2011). Until the recent passage of the America Invents Act, the American novelty rule had some nuances arising from the fact that we had a first-to-invent, rather than a first-to-file, system. See Leahy-Smith America Invents Act, Pub. L. No. 112-29, § 3, 125 Stat. 284 (2011) (amending novelty rules to establish a first-to-file system of priority); 35 U.S.C. § 102 (2006) (novelty rules that prevailed prior to passage of Leahy-Smith America Invents Act). Although the old novelty rule will continue to apply to some patents during the transition period accompanying the passage of the America Invents Act, see Leahy-Smith America Invents Act, § 3(n) (setting effective date 18 months after enactment), I set aside its nuances because they are irrelevant to understanding the purpose of the novelty doctrine at the level of generality that I discuss it.

22. Rebecca S. Eisenberg, *Analyze This: A Law and Economics Agenda for the Patent System*, 53 VAND. L. REV. 2081, 2088 (2000). This notion that the patentee must provide something “new” in order to obtain the right to exclude has been long recognized. See, e.g., 1 WILLIAM C. ROBINSON, THE LAW OF PATENTS FOR USEFUL INVENTIONS § 221 (Little, Brown, and Co. 1890).

23. 35 U.S.C. § 103; *KSR Int'l Co. v. Teleflex Inc.*, 550 U.S. 398, 417 (2007); *Graham v. John Deere Co.*, 383 U.S. 1, 14 (1966).

24. *Merges*, *supra* note 1, at 592 n.41.

25. 35 U.S.C. § 112 (2006). I have left out the written description requirement. Insofar as this rule requires anything other than what is already part of the enablement, best mode, and definiteness rules, it is to make it cheaper for the patent system to evaluate whether the invention has met the other patentability requirements. As the Federal Circuit recently put it when it held that section 112 includes a written description requirement distinct from enablement: “[a] description of the claimed invention allows the [PTO] to examine applications effectively; courts to understand the invention, determine compliance with the statute, and to construe the claims; and the public to understand and improve upon the invention and to avoid the claimed boundaries of the patentee’s exclusive rights.” *Ariad Pharms., Inc. v. Eli Lilly & Co.*, 598 F.3d 1336, 1345 (Fed. Cir. 2010) (en banc). Note that the last two goals are the same as those served by the enablement (“understand and improve upon the invention”) and definiteness (“avoid the claimed boundaries”) rules. *Id.* In short, the written description’s only unique purpose is to reduce administrative costs. Mistakes regarding compliance with the written description requirement may thus affect the cost of assessing compliance with other patentability rules

inventor to reveal important technical information regarding the invention. Enablement requires that the patent's specification teach someone knowledgeable in the field how to "make and use" the invention.²⁶ The best mode requirement demands that the inventor reveal her preferred way of making and using it.²⁷ These doctrines lighten the informational burden borne by those who seek to build on the inventor's contribution.²⁸

Finally, there is the definiteness doctrine, which requires that the inventor describe the invention in clear and precise terms.²⁹ This rule is intended to provide notice as to the legal bounds of the inventor's rights to exclude.³⁰ That notice makes it easier for third parties to avoid infringement and for the inventor herself to sell or transfer her rights.

C. Mapping the Terrain of Patent Mistakes

The patentability rules are shortcuts. If we had unlimited resources and perfect information, it would be possible simply to determine whether granting a given patent would, on net, promote the goals of patent law. Put differently, an omniscient actor would have no use for the patentability rules because patent law's goals provide sufficient grounds for deciding whether to grant a patent. The actor would grant only those patents that pass a cost-benefit analysis incorporating whatever criteria are made relevant by the goals of patent law. When we ask, for example, whether an invention is novel and non-obvious, we are in effect asking something like whether the invention is the kind of thing for which the prospect of a patent provides incentives that justify the deadweight losses of the patent grant. With unlimited resources and perfect information, an omniscient actor could answer that latter question directly, rather than rely on the indirect answer to it provided by the analysis of whether the invention is novel and non-obvious.³¹ Simply stated, the patentability

and may therefore affect the rate of errors. But it will not affect the costs of the errors that we do make. Because this Article's focus is on the costs of errors, rather than their rate, I do not discuss written description separately from the other disclosure rules.

26. 35 U.S.C. § 112.

27. *Id.*; *Eli Lilly*, 251 F.3d at 963 (requiring both a subjective inquiry into the inventor's own view of the best mode of practicing the invention and an objective inquiry into the sufficiency of the disclosure of the best mode found in the specification).

28. See Matthew H. Solomon, *Patently Confusing: The Federal Circuit's Inconsistent Treatment of Claim Scope as a Limit on the Best Mode Disclosure Requirement*, 45 IDEA 383, 385 (2005) (noting that best mode requirement is designed to place inventor and rivals on equal footing).

29. 35 U.S.C. § 112.

30. *All Dental Prodx, LLC v. Advantage Dental Prods., Inc.*, 309 F.3d 774, 779-80 (Fed. Cir. 2002) ("The primary purpose of the definiteness requirement is to ensure that the claims are written in such a way that they give notice to the public of the extent of the legal protection afforded by the patent.").

31. As a result, it is plausible to conceptualize even perfect applications of the patentability rules as mistakes—sometimes an invention that is novel and non-obvious will

rules are surely overinclusive and underinclusive with respect to their underlying purposes.

An omniscient actor would therefore have no use for the patentability rules. Nonetheless, for the sake of conceptual clarity, I will assume that the patentability rules were designed to be applied by an omniscient actor, who could always determine whether a given invention was useful, novel, and so on. Whether because of institutional design,³² cognitive constraints,³³ or imperfect information,³⁴ the patent system will sometimes reach conclusions that differ from the ones that the omniscient actor would reach. Mistakes thus occur not because the rules themselves align imperfectly with their underlying purposes, but rather because we cannot know to a certainty whether a given invention complies with the requirements of the rules. For purposes of this Article, then, a mistake occurs whenever the patent system's application of the patentability rules reaches a decision (grant or deny) at odds with the decision that an omniscient actor applying the rules would reach.

Mistakes, so defined, will inevitably occur when the patent system applies the patentability rules in specific cases.³⁵ Even as defined here, though, mistakes are not necessarily undesirable. It costs something to avoid a mistake, and our resources are limited. Whenever the costs of avoiding the mistake are greater than the costs of making it, we will be better off making the mistake. In short, the optimal level of mistakes is greater than zero.³⁶

In order to make optimal mistakes, we need to know both the costs of the mistake and the costs of mistake-avoidance. A complete theory of optimal mistakes is beyond the scope of this Article. I set

nonetheless be a net loss in light of the patent system's goals. An example here might be an accidental invention that was achieved without the patent system's incentives; think of the use of penicillin to treat infections. The deadweight losses incurred by granting a patent on that kind of invention may outweigh the benefits of, for example, faster distribution of knowledge.

32. Jonathan S. Masur, *Costly Screens and Patent Examination*, 2 J. LEGAL ANALYSIS 687, ¶¶ 16-24 (2010) (describing incentives that lead patent examiners to issue patents that do not meet the patentability requirements and citing the PTO's self-described mission as "to help our customers get patents"); Merges, *supra* note 1, at 609 ("The current bonus system [for PTO examiners] is believed to skew incentives in favor of granting patents.").

33. See Gregory N. Mandel, *Patently Non-Obvious: Empirical Demonstration that the Hindsight Bias Renders Patent Decisions Irrational*, 67 OHIO ST. L.J. 1391 (2006) (reporting results of experimental study indicating that hindsight bias causes systematic errors in the application of patent doctrine); see also Gregory Mandel, *Patently Non-Obvious II: Experimental Study on the Hindsight Issue Before the Supreme Court in KSR v. Teleflex*, 9 YALE J.L. & TECH. 1 (2006) (discussing a follow-up experiment).

34. Thomas, *supra* note 1, at 313 (describing broad scope of prior art that may be relevant to a patent's validity under 35 U.S.C. § 102 (2006)).

35. See Lichtman & Lemley, *supra* note 1, at 61 (contending that resource and information constraints make mistakes inevitable); Meurer, *supra* note 5, at 679 ("It is vital to recognize that examiners will make mistakes given the time constraints that they face.").

36. See generally Lemley, *supra* note 3.

aside entirely the costs of mistake-avoidance. And I do not intend to provide a precise accounting of the costs of a given mistake. The goal for now is to take an important step in the direction of optimal mistakes by explaining how the relative costs of patent mistakes vary by three important characteristics. The first is the type of mistake—false positive or false negative. The second is the timing of the mistake—whether it is made early during the application process or late during the infringement litigation process. The last is the doctrinal basis for the mistake—which of the patentability rules has the patent system wrongly applied. This Section describes these characteristics in detail.

1. *Type*

The patent system's mistakes will be one of two familiar types: false positive or false negative. A false positive mistake occurs when the patent system grants a patent on an application that does not comply with the patentability rules. Suppose someone applies today for a patent that claims a device with "two wheels, the one directly in front of the other, combined with a mechanism for driving the wheels, and an arrangement for guiding; which arrangement also enables the rider to balance himself upon the two wheels."³⁷ A bicycle. Because that exact invention was disclosed long ago, such an application would fall short of the novelty requirement.³⁸ If the patent system nonetheless grants a patent on this application, the mistake will be a false positive.

A false negative mistake occurs when an application is not granted even though it complies with the patentability rules. Suppose someone applies today for a patent on a new treatment that prevents memory loss in patients with Alzheimer's disease. Assume that the application complied with all the patentability rules, but the patent system wrongly concludes that the application does not tell a person skilled in the treatment of Alzheimer's how to make and use the invention—that is, the patent system wrongly concludes that the patent does not comply with the enablement requirement.³⁹ If no patent issues from this application, the mistake will be a false negative.

2. *Timing*

We have so far considered the government's evaluation of patentability at a high level of abstraction. Let's add some procedural details.

37. U.S. Patent No. 59,915 (filed April 1866).

38. See 35 U.S.C. § 102 (2006); *supra* notes 20-24 and accompanying text.

39. See 35 U.S.C. § 112 (2006); *supra* notes 25-28 and accompanying text.

The government may evaluate patentability twice.⁴⁰ I call the first evaluation “early” and the second one “late.” The terms early and late are relative. I use them here to refer to when in the life of the patent the evaluations occur.

Begin with the early evaluation. Suppose an inventor finds a way to make a paper cup that can insulate beverages much better than existing paper cups. It can be used equally well for drinking freshly-brewed coffee or ice-cold beer. She decides to obtain a patent on her insulating paper cup.

In order to do so, she must first submit an application to the Patent and Trademark Office (PTO),⁴¹ which assigns it to an examiner.⁴² If the examiner decides that the application complies with all of the patentability rules, he issues a patent.⁴³ If not, he must tell the inventor why he thinks the application is deficient,⁴⁴ and the inventor has the opportunity to respond.⁴⁵ This process may be repeated until (1) the examiner grants a patent on the application; (2) the applicant

40. I am setting aside intermediate examinations that occur when administrative proceedings are invoked after the patent has issued. See 35 U.S.C. §§ 301-307 (2006) (establishing *ex parte* procedures for evaluating the patentability of issued patents); 35 U.S.C. §§ 311-318 (2006) (establishing *inter partes* procedures for evaluating the patentability of issued patents) (amended by Leahy-Smith America Invents Act, Pub. L. No. 112-29, § 6(a), 125 Stat. 299 (2011) (prohibiting the institution of *inter partes* reexaminations until after a post-grant opposition window closes, increasing the threshold for instituting *inter partes* reexaminations, and providing that such reexaminations will be conducted by three-judge panels of the PTO)); Leahy-Smith America Invents Act § 6(d) (creating a post-grant review proceeding that can be instituted in the first nine-months after the patent issues). I set these proceedings aside because they might occur at any point in time—shortly after issuance or just before (or even during or after) litigation. So there is little to be said generally about the timing of administrative patentability review. To the extent that the review occurs close in time to the PTO’s initial decision to issue a patent, the review might be said to be early and the analysis of early mistakes applies. (Note that this will be true of all proceedings under Leahy-Smith America Invents Act, § 6(d).) To the extent that the review occurs long after the PTO issues the patent, the review might be said to be late and the analysis of late mistakes applies.

41. 35 U.S.C. §§ 111-113 (2006).

42. 4-11 DONALD S. CHISUM, CHISUM ON PATENTS § 11.01 (2011).

43. 35 U.S.C. § 131 (2006).

44. 35 U.S.C. § 132 (2006) (requiring that the examiner “stat[e] the reasons for [a] rejection”). The examiner’s rejections should be complete to the extent possible; that is, they must identify the fundamental defects in the application, although there is some leeway for alternative denials on novelty and nonobviousness grounds, as well as for instances in which the indefiniteness of the claims prevents the examiner from comparing them to the prior art. 4-11 CHISUM, *supra* note 42, § 11.03(1)(c)(i).

45. The inventor can demand that the examiner conduct at least one reexamination of the application. 35 U.S.C. § 132. The inventor’s response might explain why she disagrees with the examiner (including submitting evidence and affidavits to demonstrate patentability), amend or cancel problematic claims, or modify the specification. 4-11 CHISUM, *supra* note 42, § 11.03(2)(a)(i). The response cannot, however, introduce “new matter” into the application. 35 U.S.C. § 132. In order to introduce new matter, the applicant must submit what is known as a continuation-in-part to the application, which applies a later filing date to the newly-added material. 4A-13 CHISUM, *supra* note 42, § 13.03(3).

abandons the application; or (3) the applicant appeals the examiner's rejection and a final decision issues from the courts.⁴⁶ The early decision is made whenever any of these three events occur—grant, abandonment, or a judicial decision on the application's patentability. If the outcome here deviates from what would have occurred if an omniscient actor had evaluated the application's compliance with the patentability rules, there has been an early mistake.

Now turn to the late evaluation. Suppose the PTO had granted the inventor a patent on her insulating paper cup. Some months or years later, a rival begins making and selling paper cups incorporating the patented technology.⁴⁷ The patentee may sue her rival for infringement.⁴⁸ In such a lawsuit, the rival may defend himself on the ground that the patent is invalid.⁴⁹ If he does so, the patent system—here the court or jury—must again evaluate whether the patent complies with the patentability rules. If it concludes that the patent does not comply with the patentability rules, the patent is invalidated.⁵⁰ If it concludes that the patent does comply with the patentability rules, the challenge is rejected. If the outcome here deviates from what would have occurred with an omniscient actor, there has been a late mistake.

Note that a late decision could only occur if the patent system had issued a patent at the early stage, so any late false positive must therefore have been preceded by an early false positive. If, at the ear-

46. Technically, the applicant would first have to seek review within the PTO. See 35 U.S.C. § 134(a) (2006) (permitting appeal to the Board of Patent Appeals and Interferences ("BPAI") by "[a]n applicant for a patent, any of whose claims has been twice rejected"); 35 U.S.C. § 141 (2006) (permitting appeal to the Federal Circuit from decisions of the BPAI); 35 U.S.C. § 145 (2006), *amended by Leahy-Smith American Invents Act*, Pub. L. No. 112-29, § 9(a), 125 Stat. 284, 316 (2011) (creating cause of action in district court to seek issuance of patent after BPAI rejection).

47. Since Congress established the Federal Circuit, the average litigated patent is six years old from the date of issuance to the date that litigation begins and nine years from issuance to the date of decision. Scott E. Atkinson, Alan C. Marco & John L. Turner, *The Economics of a Centralized Judiciary: Uniformity, Forum Shopping, and the Federal Circuit*, 52 J.L. & ECON. 411, 431 (2009).

48. 35 U.S.C. § 271(a)-(b) (2006) (allowing the patentee to sue for infringement anyone who "without authority makes, uses, offers to sell, or sells" the invention). Only a small percentage of issued patents in fact reach litigation. See, e.g., Lemley, *supra* note 3, at 1501 (over 98% of patents are never litigated). Late evaluations by the patent system are the exception, not the rule; for most patents, the patent system typically only conducts an early evaluation. But more valuable patents are more likely to be litigated and, therefore, to receive a second, late evaluation. See John R. Allison, Mark A. Lemley, Kimberly A. Moore & R. Derek Trunkey, *Valuable Patents*, 92 GEO. L.J. 435, 439-43 (2004) (explaining why "litigated patents tend to be much more valuable than others on average" and why "valuable patents are much more likely than others to be litigated").

49. 35 U.S.C.A. § 282 (2011).

50. That evaluation accords significant weight to the early decision to grant the patent. See *Microsoft Corp. v. i4i Ltd. P'ship*, 131 S. Ct. 2238 (2011) (holding that a party who asserts that a patent is invalid must present clear and convincing evidence of invalidity). But set this nuance aside for the moment. The important point is that the government again evaluates whether the inventor has complied with the patentability requirements.

ly stage, the patent system denied the application (rightly or wrongly), there will be no late decision—there cannot be infringement litigation if no patent has issued. And if, at the early stage, the patent system correctly granted the application, then a late decision affirming the patent’s validity will be a true positive, by definition. Only if the early decision was an incorrect grant of an application—that is, an early false positive—might the patent system commit a late false positive at all.

Patent mistakes may therefore be made early or late. An early evaluation occurs at the patent’s birth. Ordinarily, a late evaluation occurs when the patent is well into adulthood.⁵¹ Though I use the terms early and late to link the patent system’s decisions (and its mistakes) to the lifespan of the patent, those decisions might also correlate with important events outside the life of the patent, like product launches. These correlations will vary across industries. In the pharmaceutical industry, for example, an early evaluation will occur before the firm launches a drug incorporating the patented invention, and a late evaluation will occur after the drug is on the market. In the software industry, on the other hand, both early and late evaluations often occur after product launch. The industry-specific correlation of timing within the patent system to timing outside the patent system will affect, on an industry-specific basis, the relative assessment of mistakes made at different times.

3. *Doctrine*

As the above examples suggest, mistakes can be made on any of the doctrinal bases. The bicycle patent was wrongly issued despite its failure to comply with the novelty requirement. The Alzheimer’s patent was wrongly denied for failure to comply with the enablement requirement. And, of course, we could imagine mistakes on any of the patentability rules.

The application of the patentability rules—at the early and late stages—is all-or-nothing.⁵² If the patent system concludes at the early stage that any one of the rules is violated, it will not grant a patent on the application. And if the patent system concludes at the late stage that any one of the rules is violated, it will invalidate the patent. Only when the patent system concludes that all of the patentability rules are satisfied will it grant a patent (at the early stage) or uphold the validity of an issued patent (at the late stage).

51. See Atkinson et al., *supra* note 47, at 431.

52. See *Festo Corp. v. Shoketsu Kinzoku Kogyo Kabushiki Co.*, 535 U.S. 722, 736 (2002) (listing the “statutory requirements [that] must be satisfied before a patent can issue” and noting that the “failure to meet these requirements could lead to the [wrongly] issued patent being held invalid in later litigation”).

III. HOW TO MAKE BETTER MISTAKES

The last Part identified the key characteristics of the mistakes that the patent system might produce. I have assumed (and will continue to assume) that if the patent system could apply the rules without cost and without ever making a mistake, the patentability rules would produce a first-best world. But we are in a second-best world in which the patent system will make mistakes, for any number of reasons.⁵³

This Part will try to make progress by providing a sense of the relative costs of patent mistakes. Each mistake produces some negative consequences—some mistakes might undermine incentives to innovate, others might make it too hard for the public to access information, and others might generate too much legal ambiguity. These potential deviations from ideal outcomes, however, are not the end of the story. Instead, private parties will respond to patent mistakes.⁵⁴ The degree to which they can do so will depend on each mistake's type, timing, and doctrinal basis. The aim here is to identify the most likely scenarios regarding how the costs of mistakes vary and to make explicit the assumptions that would lead us to favor alternative views.

A. *False Positives*

We begin with false positive mistakes. This Section argues that because each doctrine is designed to prevent a particular kind of problem, and the kind of problem posed by a wrongly-issued patent thus turns on which doctrine it violated, the costs of false positives will vary by doctrine. Furthermore, timing has doctrine-specific effects on the costs of false positives. Because any late false positive must have been preceded by an early false positive, the relevant question is whether the late false positive imposes costs in addition to what we have already incurred as a result of the early false positive. For the definiteness and disclosure rules, the surprising answer is no—the problems posed by early false positives on these rules either (1) will be resolved by the late mistake (in the case of definiteness); or (2) will be much less serious by the time we get to the late decision (in the case of disclosure). For the scope and invention rules, however, the answer is yes—the problems posed by early false positives on these rules will likely persist after late false positives. It is therefore less important to avoid late false positives on definiteness and disclosure than for scope and invention. And, if we want to avoid

53. See *supra* text accompanying notes 35-36.

54. Indeed, “[o]ur patent system envisions a mixture of public *and* private expenditures to determine the validity of patents.” Merges, *supra* note 1, at 596.

the costs of false positives on definiteness and disclosure, we will have to do so early.

1. *Doctrine as a Direct Selection Mechanism*

As previously noted, the patentability rules can be (roughly) sorted into four categories, based on the goals served by each set of rules. The scope rules limit the inventor's ability to control downstream innovation.⁵⁵ The invention rules reject patents on inventions that society has either already obtained or will likely obtain in short order.⁵⁶ The disclosure rules make it easier for third parties to improve on or design cheap substitutes for the invention.⁵⁷ And the definiteness rules help the inventor, her rivals, and third parties organize their affairs without inadvertently exposing themselves to legal liability.⁵⁸

These goals might reasonably be contested. We might also debate whether the patentability rules as currently designed are the most effective means for achieving the goals. Maybe we cannot be certain what, exactly, the novelty doctrine does and whether it does so as well as it could.⁵⁹ And maybe there is some overlap in the doctrines—novelty and nonobviousness seem to pursue related ends, as do enablement and best mode.⁶⁰ But we can at least be confident that, whatever it is that the novelty rule does, it is something different than what the enablement rule does. And the general categories into which I have placed the patentability rules seem plausibly correct. In any event, the relevance of doctrine for false positives does not depend on these precise goals being right or this precise categorization of the doctrines being accurate. Instead, there is only one condition that must be true: The patentability rules must not all do exactly the same thing.⁶¹ That much, at least, seems right.

55. See *supra* text accompanying notes 16-19.

56. See *supra* text accompanying notes 20-24.

57. See *supra* text accompanying notes 25-28.

58. See *supra* text accompanying notes 29-30.

59. I have suggested that novelty prevents the issuance of patents that remove information from the public domain, rather than add to it. See *supra* text accompanying note 22. Some aspects of the novelty requirement—for example, those relating to whether references that are difficult to find qualify as prior art—suggest that the rule is designed to promote efficient choices between independent research and searches of existing knowledge. ROBERT P. MERGES & JOHN F. DUFFY, *PATENT LAW & POLICY: CASES AND MATERIALS* 419-23 (3d ed. 2002).

60. For example, it's no stretch to think that novelty does something like what nonobviousness does, but that applying the novelty rule first lowers the rate of errors on the nonobviousness rule. MERGES & DUFFY, *supra* note 59, at 372-73. This view would emphasize that novelty requires a precise identification of the differences between the alleged invention and the prior art, while nonobviousness requires an assessment of the magnitude of those differences. *Id.* That latter, more complex assessment will likely be more accurate if conducted separately from the former. *Id.*

61. The categories and theoretical explanations I use can be easily replaced with any alternative explanations and the same conclusions will hold. Of course, there would have to

The doctrinal basis of a false positive affects its costs by selecting for patents that pose different problems. Each patentability rule is designed to prevent or mitigate a problem that would arise if the rule were not enforced. When the patent system mistakenly does not enforce any one of the patentability rules, we bear the costs of the problems that rule is designed to mitigate. This is a necessary consequence of differentiated patentability rules.

As an example of how this works, suppose that the patent system commits false positive mistakes on two separate patent applications.⁶² The first application complies with all of the patentability rules except for definiteness. The patent system nonetheless issues a patent because it fails to recognize that the application violates that rule; it thus commits a false positive on definiteness. The second application complies with all of the patentability rules except for enablement. The patent system, however, fails to recognize that the application violates that rule; it thus commits a false positive on enablement and wrongly issues a patent on the second application.

The patent system committed a false positive on each application, and we have two patents that we'd be better off not having. But although we'd be better off without either of these two patents, they differ in important respects, and those differences are a necessary consequence of the doctrinal bases for the mistakes. Because the patent system committed a false positive on definiteness for the first patent, we know that it does not clearly define the scope of the patentee's rights.⁶³ The second patent does. And because the patent system committed a false positive on enablement for the second patent, we know that it does not tell someone working in the field how to make and use the invention it covers.⁶⁴ The first patent does. These patents thus pose distinct problems.

It is difficult to assess whether false positives on enablement are costlier than, for example, false positives on definiteness. We should, however, expect the costs to differ by doctrine. Given the differences in the types of problems caused by false positives on different doctrines, the costs should differ too. This doctrine-by-doctrine variation

be tweaks to account for how alternate categorizations or explanations depart from the standard view I adopt here. So if we think that novelty is primarily about efficient search, rather than avoiding patents that remove information from the public domain, then false positives on novelty will encourage waste of resources developing inventions that could be more cheaply obtained by reviewing existing literature. This is a different kind of problem than the ones I set out in the text, and it will impose its own costs. But the main argument is simply that the costs of false positives will vary by doctrine, and this is true whether novelty is about promoting efficient search or avoiding patent thickets.

62. I assume here for simplicity's sake that the patent system makes only one mistake per patent. The analysis can be easily extended to multiple mistakes. *See, e.g., infra* notes 104-05.

63. *See* 35 U.S.C. § 112 (2006).

64. *See id.*

in the costs of false positives is a necessary feature of any patent system that includes differentiated patentability rules, as ours does.

2. *Timing Effects*

The effects of timing on the costs of false positives depend on their doctrinal bases. This subsection thus assesses the effect of timing on the costs of false positives by doctrinal category. Because late false positives are necessarily preceded by early false positives,⁶⁵ we will want to know the additional costs imposed by a late false positive so that we can think about whether it's worth risking a late false negative in order to avoid a late false positive. If a late false positive on some doctrine is roughly indistinguishable from a late true negative, it's unlikely that it will be worth the resources spent getting it right or the risk of getting it wrong by wrongly invalidating a patent that should be upheld. This analysis can also tell us whether the early decision is our only chance to avoid the costs of the mistake. It would be a poor strategy to count on litigation to fix examination errors on doctrines for which we will incur most of the costs before litigation.

As we will see, for the scope and invention rules, a late false positive might impose significant additional costs. For the disclosure and definiteness rules, however, the additional costs incurred as a result of a late false positive will likely be low. We should therefore be more willing to tolerate late false positives on the latter rules than on the former.

(a) *Disclosure Rules*

The disclosure rules—enablement and best mode—attempt to ensure that patents communicate information about the invention.⁶⁶ False positive mistakes on these rules permit inventors to obtain and enforce patents even when those patents do not convey important information to those who read them. The problem caused by these mistakes is that the inventor's rivals will have to do more of their own research to copy the invention than the patentability rules deem optimal.⁶⁷ Disclosure false positives force a wasteful duplication of resources by rivals trying to accomplish what the inventor has already done.

65. See *supra* text accompanying notes 50-51.

66. 35 U.S.C. § 112. Again, I set aside written description because its only independent purpose is to reduce the administrative costs involved in assessing whether the other patentability rules are met. See *supra* note 25.

67. There is a long-standing debate in the patent literature regarding the extent to which duplication of research efforts—known as patent races—are good or bad. For introductions to this debate, see Michael Abramowicz, *The Uneasy Case for Patent Races over Auctions*, 60 STAN. L. REV. 803 (2007) and Tim Wu, *Essay, Intellectual Property, Innovation, and Decentralized Decisions*, 92 VA. L. REV. 123 (2006).

There is an important difference between early and late false positives for these rules. Though the inventor found a solution to a technical problem, it may not be the only one. With the passage of time, other solutions may present themselves, and rivals will be able to build on the technical information revealed by them. To the extent that enablement and best mode try to put the inventor and her rivals on equal footing at the moment of patent issuance, that goal is more crucial when the invention is first patented than during infringement litigation when other advances may have already achieved the same effect. And assuming that the infringement allegations are not frivolous, the fact of litigation indicates that the defendant has managed to duplicate the inventor's achievement (or at least come close), be it by copying, reverse engineering, independent invention, or otherwise. In short, late false positives on disclosure occur after the costs of the mistake have been incurred. Because the harm from inadequate disclosure comes early in the patent's life, late false positives contribute little to the overall costs of the mistake. If so, then we need not worry much about late disclosure false positives, at least as compared to early ones.

(b) *Definiteness*

Definiteness is designed to ensure that the patent's claims define the boundaries of the patentee's legal rights to exclude others from the invention.⁶⁸ When the patent system commits a false positive on definiteness, third parties will be unsure whether their activities expose them to potential infringement liability.⁶⁹ And transactions involving patents that do not comply with the definiteness doctrine will be more expensive as the licensee cannot be sure what, precisely, she is buying.

There is little that private parties can do to fix definiteness false positives. Perhaps they can pay more in legal fees to try to understand the meaning of poorly-defined claim terms. Of course, to the extent that definiteness false positives make patent boundaries difficult to ascertain, they may also create something like a patent thicket—there may be many patents in a field with unclear boundaries, and those who wish to participate in the field will have trouble navigating a path to liability-free activities.⁷⁰ In those instances, cross-licensing and patent pools might help.

68. 3-8 CHISUM, *supra* note 42, § 8.03 (“The primary purpose of this requirement of definiteness in claims is to provide clear warning to others as to what constitutes infringement of the patent.”).

69. See *All Dental Prodx, LLC v. Advantage Dental Prods., Inc.*, 309 F.3d 774, 779-80 (Fed. Cir. 2002).

70. See notes 85-90 and accompanying text.

As with the disclosure rules, there is an important difference between definiteness false positive mistakes made early at the application stage and those made late at the litigation stage. The definiteness rule holds only that claims that are “‘not amenable to construction’ or ‘insolubly ambiguous’ are indefinite.”⁷¹ When the patent system commits an early false positive on definiteness, the issued patent simply stands with incomprehensible claims. But when the patent system commits a late false positive on definiteness, it does so in the context of litigation that includes a *Markman* hearing.⁷² In a *Markman* hearing, the court construes the claim and thus supplies some comprehensible meaning where previously none existed. Assuming that false positives on definiteness will occur, the ones that occur during litigation are less troubling because they solve the problem that the definiteness doctrine tries to avoid. Perhaps some residual ambiguity remains, but late false positives provide a definite meaning to at least some previously incomprehensible part of the patent’s claims.

Late mistakes on definiteness therefore inform third parties of the scope of the patentee’s right to exclude. Put differently, late false positives at least reduce the uncertainty that makes definiteness false positives problematic in the first place. Because late false positives on definiteness impose little additional cost to that already incurred as a result of the early false positive, the patent system might therefore do well to focus its late resources on other doctrines.

(c) *Scope Rules*

The scope rules include the subject matter and utility requirements. The former prohibits patents on laws of nature, abstract ideas, and natural phenomena.⁷³ The latter requires a demonstration that the invention is good for something other than being the subject of scientific inquiry.⁷⁴ False positives on these rules grant patentees excessive control over downstream innovation.⁷⁵ The danger is that patents violating these rules will frequently lead to blocking-patent dynamics that can hinder technological progress.⁷⁶

Blocking patents occur when one invention builds on a prior one; both inventions are patented, but the patent for the second invention

71. *Datamize, LLC v. Plumtree Software, Inc.*, 417 F.3d 1342, 1347 (Fed. Cir. 2005).

72. The hearings are named after *Markman v. Westview Instruments, Inc.*, 517 U.S. 370 (1996). That case held that the interpretation of patent claims is a legal question to be decided by the court, rather than a question for the jury. *Id.* at 391.

73. See *supra* note 17 and accompanying text.

74. See *supra* note 19 and accompanying text.

75. See *supra* note 18 and accompanying text.

76. Michael Risch, *Reinventing Usefulness*, 2010 BYU L. REV. 1195, 1224 (“One normative reason to deny patentability to compositions of unknown use is to avoid inefficient blocking patents.”).

falls within the scope of the claims covered by the patent for the first invention.⁷⁷ This situation might arise when, for example, one inventor patents a chemical compound, and a second inventor discovers and patents a non-obvious use for that compound.⁷⁸ The initial inventor can freely produce the compound but cannot practice the newly-discovered use without the second inventor's permission.⁷⁹ And because the second inventor cannot produce the compound without the first inventor's permission, she cannot practice her newly-discovered use unless and until the first inventor allows her to do so.⁸⁰ This situation may lead to a bargaining breakdown because it is difficult to allocate the value of each party's contribution to the combination.⁸¹ In these circumstances, the parties may reach (or at least assert) widely divergent valuations for their relative contributions and therefore may be unable to come to an agreement on how to split the surplus from combining their patents.⁸²

Utility false positives will frequently lead to these dynamics because they grant patents when there is no known use for the invention.⁸³ Any productive use will be discovered later, either by the inventor or a third party. Subject matter false positives—which permit inventors to obtain patents on abstract ideas, laws of nature, or natural phenomena—will similarly grant patents that confer on their holders exclusive rights over broad areas.⁸⁴ Because second-comers will frequently encroach on those broad areas, these false positive mistakes will also tend to produce blocking patent scenarios.

Early false positives on the scope rules will be costly. Some inventors may avoid an area if they fear getting caught in a blocking dynamic because of an early false positive, and bargaining breakdown may occur even if the patent is never asserted in infringement litigation. And late false positives likely impose significant costs in addition to those imposed by the early false positive. The solution to a blocking patent dynamic is bargaining, and there is little reason to think that bargaining will be easier if there has been a late false positive than if there has only been an early false positive. Perhaps the passage of time reveals better information about the value of each

77. Robert Merges, *Intellectual Property Rights and Bargaining Breakdown: The Case of Blocking Patents*, 62 TENN. L. REV. 75, 80-81 (1994).

78. Risch, *supra* note 76, at 1224.

79. This is because patents confer only a right to exclude others, not an affirmative right to use. See 35 U.S.C.A. § 271 (West 2012).

80. *Id.*

81. Merges, *supra* note 77, at 89 (“Where high uncertainty attends the valuation of assets to be exchanged, bargaining can be difficult.”).

82. *Id.*

83. See *Brenner v. Manson*, 383 U.S. 519, 529 (1966). See *supra* note 19 and accompanying text.

84. *Diamond v. Chakrabarty*, 447 U.S. 303, 309; *accord Bilski v. Kappos*, 130 S. Ct. 3218, 3225 (2010). See *supra* note 17 and accompanying text.

party's contribution so that bargaining is easier later. But it is also plausible that the passage of time will merely prolong the bargaining stalemate. As a result, we cannot dismiss late false positives on these rules as unimportant; both early and late false positives here are costly.

(d) *Invention Rules*

The invention rules include the novelty and nonobviousness requirements.⁸⁵ When the patent system commits false positives on these rules, it risks the creation of patent thickets—areas of technology in which many parties can assert overlapping rights to any given invention.⁸⁶ In a patent thicket, a license from a single patentee is useless because another rights-holder can veto the permission granted by the first licensor.⁸⁷ This dynamic increases both the number of transactions inventors must complete and the risk of bargaining breakdown.⁸⁸

Although thickets increase the likelihood of bargaining breakdown, they do not make it inevitable. Instead, there are two potential solutions to patent thickets: cross-licensing agreements and patent pools. When two firms each hold patents that the other infringes, they can enter into a cross-licensing agreement—that is, they agree to grant licenses that permit each to use the other's patents.⁸⁹ They can also be extended to include more than two parties. Patent pools are similar to cross-licenses in that they also bundle patents together to cut through a patent thicket's overlapping rights to exclude. The

85. 35 U.S.C. §§ 102, 103 (2006).

86. Patent thickets are a version of the anticommons problem. Anticommons occur whenever property rights are allocated such that many parties have the right to exclude others from a resource and no single party has the right to use it. See Burk & Lemley, *supra* note 9, at 1611 (“[A] pure anticommons involves . . . different contributions that must be aggregated together”); Heller & Eisenberg, *supra* note 1, at 698 (describing the tragedy of the anticommons as a situation in which “multiple owners each have a right to exclude others from a scarce resource and no one has an effective privilege of use”). Anticommons can arise without the overlap in rights created by invention false positives; all that is needed for an anticommons is that rights must be aggregated in order to use a resource. This may occur in the patent context when, for example, one invention incorporates several other invention, as often happens in the auto industry. See Robert P. Merges, *Intellectual Property Rights and the New Institutional Economics*, 53 VAND. L. REV. 1857, 1859 (2000) (identifying cars and consumer electronics as “multi-component products” for which an “individual patent[] often cover[s] . . . a single component or sub-component”). Insofar as we are concerned with invention false positives, though, we are concerned with overlapping rights, not simply rights that must be combined to be useful. So, we can restrict our focus to the patent thicket. See Burk & Lemley, *supra* note 9, at 1627 (“Anticommons exist where several different inputs must be aggregated together to make an integrated product. Patent thickets, by contrast, occur when multiple intellectual property rights cover the same technology and therefore overlap.”).

87. See *Standard Oil Co. v. United States*, 283 U.S. 163, 171 (1931).

88. See generally Carl Shapiro, *Navigating the Patent Thicket: Cross Licenses, Patent Pools, and Standard Setting*, in 1 INNOVATION POLICY AND THE ECONOMY 119 (Adam B. Jaffe et al. eds., 2001) (describing the bargaining dynamic in a patent thicket).

89. *Id.* at 127.

difference is that pools make the bundled patents available to third parties who have no patents of their own to add to the bundle.⁹⁰

Here, again, the additional costs of a late false positive might be high. There is no general reason to think that parties will be more or less likely to enter into cross-licensing arrangements or patent pools after a late false positive on the invention rules. So to the extent that early false positives on the invention rules create thickets, those problems are likely to persist after late false positives.

3. *Summary and Caveats*

False positives on different doctrines thus create different kinds of problems. It is therefore likely that the costs of false positives vary by doctrine too—it would be an unusual coincidence if the costs of patent thickets created by false positives on novelty and nonobviousness happened to precisely match the costs of wasteful duplication of resources necessitated by false positives on enablement and best mode.

The relationship between timing and the costs of false positives also varies by doctrine. Even after late false positives on the scope and invention rules, we are likely to incur significant costs; therefore, we have reason to be concerned about compliance with those rules at both the early and late stages. For the disclosure and definiteness rules, on the other hand, we will incur little additional costs after late false positives. Accordingly, we have less reason to be worried about compliance with those rules at the late stage than we do at the early stage. And, as between doctrines at the late stage, we should be more concerned about compliance with the scope and invention rules than the disclosure and definiteness ones.

Though I have noted that the costs of false positives vary by doctrine, I have not indicated which doctrines produce more or less costly false positives. Doing so would require resolving highly-contested issues about the purposes of the patent system that are beyond the scope of this Article. Consider the persistent debates between prospect theorists and reward theorists.⁹¹ Prospect theorists contend that the patent system ought to grant broad rights at the earliest possible time, to encourage the holders of those rights to spend resources developing them and to facilitate licensing.⁹² Reward theorists, on the

90. *Id.* Of course, cross-licensing agreements and patent pools are costly to form for all the reasons that thickets are problematic in the first place. They might also raise antitrust concerns in some contexts. *Id.* The point is not that these mechanisms will always solve the thickets problem, only that they might.

91. See generally Mark A. Lemley, *Ex Ante versus Ex Post Justifications for Intellectual Property*, 71 U. CHI. L. REV. 129 (2004).

92. See *id.* at 132-33; Edmund W. Kitch, *The Nature and Function of the Patent System*, 20 J.L. & ECON. 265 (1977) (introducing the prospect theory by analogy to mining prospects).

other hand, argue that the patent system ought to grant the narrowest rights possible that will provide sufficient incentive for the inventor to create because narrow rights minimize deadweight losses produced by the patent monopoly.⁹³ If prospect theory is correct, and patents are designed to grant control over a broad area of potentially valuable research, then false positives on utility and subject matter might be fairly unproblematic.⁹⁴ But if reward theory is correct, and patents are designed to grant the lowest-cost monopoly that would produce an invention, then false positives on utility and subject matter may be the most troubling false positives. Similarly, it may be that disclosure is an essentially meaningless feature of the patent system, such that false positives on those doctrines are almost irrelevant.⁹⁵ Or, it may be that disclosure is an essential part of the bargain between the inventor and the public.⁹⁶ Whether prospect theory or reward theory is right and whether disclosure is a useless appendage or a key feature of the patent system will determine our assessment of which false positives are more or less costly. I leave for another day the resolution of these persistent questions in patent theory.

I have suggested that most of the costs of false positives on disclosure and definiteness will be attributable to the early decision and that little harm might come from following that early mistake with a late false positive. That conclusion must be qualified by two general considerations. First, any false positive is worrisome not only for the specific problems it causes in any given situation, but also because it will increase efforts to evade the patentability rules that are the source of the false positive. If the patent system produces a false positive on enablement, subsequent inventors will be more likely to seek patents while avoiding compliance with the enablement requirements—they can reap the benefits of the mistake (preventing rivals from accessing information that could help them design cheap copies) without bearing the costs (wasteful duplication of resources as many people work to achieve the same result). The costs of false positives therefore include not only the direct costs of the mistake, but also the indirect costs of increasing efforts to produce more mistakes. A late false positive will be easier for third parties to observe than an early false positive, if only because there are many more issued patents

93. See Lemley, *supra* note 91, at 131.

94. See F. Scott Kieff, *The Case for Registering Patents and the Law and Economics of Present Patent-Obtaining Rules*, 45 B.C. L. REV. 55, 105-08 (2003) (arguing that the subject matter and utility doctrines should be abolished).

95. See Alan Devlin, *The Misunderstood Function of Disclosure in Patent Law*, 23 HARV. J.L. & TECH. 401 (2010) (arguing that disclosure is ancillary to the patent system's main purpose of providing incentives to innovate).

96. See Fromer, *supra* note 9 (arguing that disclosure is of central importance to the patent system).

than litigated ones.⁹⁷ As a result, we should expect that early false positives will do little to affect the public's incentives to evade the patentability rules, while late false positives (regardless of doctrine) may significantly increase those incentives.⁹⁸

In addition, early false positives are easier for the patent system itself to fix than late false positives. Suppose an application does not comply with one of the patentability rules, but the patent system commits an early false positive and issues the patent anyway. If the patentee later seeks to enforce the wrongly-issued patent, the alleged infringer can challenge its validity, and the court therefore has the opportunity to produce the correct result—invalidate the patent. True, late false positives may also be revisited by subsequent courts.⁹⁹ But it is less likely that a court will invalidate a patent that has already withstood a validity challenge.¹⁰⁰ If we are confident in the patent system's ability to self-correct, then perhaps we should not worry much about early false positives. The widespread concern with invalid patents suggests that this ability is limited,¹⁰¹ though, and we should in fact worry about early false positives.

B. False Negatives

Turn now to false negatives. The general problem posed by false negatives does not vary by doctrine: in all cases, false negatives reduce inventors' incentives to spend money on research. Nevertheless, although the kind of problem does not vary by doctrine, the doctrinal basis for a false negative might still affect its costs by selecting for patents with importantly different characteristics. I explore two plausible characteristics—value and technological field or industry. Though we cannot rule out the possibility that the doctrines select for patents with high or low value, that possibility is not inevitable. As for industry, the patent system is likely applying the doctrines in such a way that they do select for patents in different industries;

97. Compare United States Patent and Trademark Office Patent Technology Monitoring Team, *U.S. Patent Statistics Chart Calendar Years 1963-2011*, http://www.uspto.gov/web/offices/ac/ido/oeip/taf/us_stat.htm (last modified Jan. 18, 2012) (noting that the PTO has issued over 185,000 patents per year since 2006 and that 247,713 patents were issued in 2010), with Lemley, *supra* note 3, at 1501 (noting that only about one hundred cases are litigated annually).

98. The visibility of the mistake is more important for false negatives, which generally are costly only to the extent that inventors know about them. See *infra* text accompanying notes 128-30.

99. A court's rejection of a validity challenge does not bind those who were not parties to the litigation; a court's acceptance of a validity challenge does, however, bind the patentee. See *Blonder-Tongue Labs., Inc. v. Univ. of Ill. Found.*, 402 U.S. 313, 350 (1971).

100. This is especially true when there is no new evidence to support the validity challenge; subject matter and utility might be contrasted with novelty and nonobviousness, where the discovery of new prior art may well justify a departure from a prior court's decision.

101. See *supra* note 1.

false negatives on some doctrines will therefore have industry-specific effects.

Timing also affects the costs of false negatives. Late false negatives might be less costly than early ones because they ensure that the inventor will have had at least some period of exclusivity in which to appropriate returns before the patent is (wrongly) invalidated. Still, early false negatives have two underappreciated advantages over late ones: (1) they might make it easier for inventors to turn to non-patent appropriation mechanisms that can mitigate the costs of a false negative in any given case; and (2) they might be more difficult for third parties to observe, thereby limiting the detrimental impact of false negatives on incentives to innovate.

1. *Doctrine as an Indirect Selection Mechanism*

Consider the general problem posed by false negatives. Assuming that mistake-free application of the patentability rules would optimally sort inventions that should be patented from those that should not, false negatives are costly because they prevent inventors from using patents to appropriate returns to their research investments, even when they are the best mechanism for doing so.

This view of the general problem implies that false negatives do not matter unless the inventor complied with all of the patentability rules.¹⁰² Recall that the patentability rules are all-or-nothing: if the inventor violates even one rule, the patent system denies her a patent.¹⁰³ A true negative thus renders mistakes on other rules irrelevant—the patentability rules are designed to deny an inventor a patent if she violates any one rule, and a true negative guarantees that result.

For example, suppose an inventor submits an application that does not comply with all of the patentability rules. It satisfies the novelty requirement but none of the others. If the patent system perfectly applied the patentability rules in this scenario, it would recognize that the inventor complied with the novelty requirement and violated all others. Because the patentability rules are all-or-nothing, the result is no patent. Given our baseline assumptions, this is ideal—the rules were perfectly applied, and the inventor was denied a patent.

Now take the same application and suppose that the patent system produced accurate results on all the rules, except it committed a false negative on novelty. The patent system correctly recognized

102. In order to isolate the effects of a false negative and to ease the exposition, I assume here that the patent system makes only one mistake at a time, but the analysis can easily be extended to multiple mistakes. *See, e.g., infra* note 104 (describing a hypothetical in which the patent system produces a false negative and a false positive on a single application).

103. *See Festo Corp. v. Shoketsu Kinzoku Kogyo Kabushiki Co.*, 535 U.S. 722, 736 (2002).

that the inventor violated the nonobviousness, enablement, and other rules, but it also wrongly concluded that the inventor violated the novelty rule. Again, because the patentability rules are all-or-nothing, the result is no patent. This is precisely the same result that obtained with perfect application of the patentability rules.¹⁰⁴ So the false negative on novelty made no difference. In both cases, the patent system reached the right result: it denied the inventor a patent.¹⁰⁵ Accordingly, a false negative only matters if the patent system produces true positives on all the other rules. We can therefore restrict our attention here to cases in which the inventor complies with all of the patentability rules.

Suppose an inventor submits an application that complies with all of the patentability rules—the invention is new and non-obvious, the description tells a person skilled in the art how to make and use the invention, and so on. The patent system now commits a false negative mistake on enablement—it wrongly concludes that the specification would not tell a person skilled in the art how to make and use the invention.¹⁰⁶ As a result, it denies the inventor a patent. In a perfect world, the inventor would have been able to use a patent to appropriate returns to her invention. In this imperfect world of a false negative on enablement, the inventor cannot do so. Her incentives to invest in subsequent research projects will be diminished to the extent that she thinks that the patent system might repeat its mistake.¹⁰⁷

Importantly, it would have made no difference if the false negative in the example occurred with respect to nonobviousness, written description, or any other patentability rule. Regardless of its doctrinal basis, a false negative on one rule combined with true positives on all others produces the same result: the inventor who should have been

104. Because the inventor has in fact violated all the patentability rules aside from novelty, the patent system cannot produce a false negative with respect to any rule other than novelty. Of course, if the inventor had complied with another rule—enablement, for example—the patent system could produce a false negative with respect to that rule and a true positive with respect to novelty. The result is the same as the case described in the text: the patent system rejects the application, and that's the optimal result. So this is also indistinguishable from the case in which the patent system perfectly applied the patentability rules. We could also imagine that the patent system produces a false negative on novelty and a false positive on, say, subject matter. Again, the outcome is indistinguishable from that obtained when the patent system perfectly applies the rules—the inventor is denied a patent.

105. The same conclusion would obtain regardless of which doctrine was the source of the false negative error. In any case in which the patent system produces at least one true negative (that is, any case in which the patent system correctly determines that the inventor did not comply with at least one rule) and any number of false negatives, the ideal result and the actual result would be the same: no patent.

106. 35 U.S.C. § 112 (2006).

107. And other inventors' incentives to innovate will be diminished if they are aware of the false negative. This feedback effect is explored in more detail *infra*, text accompanying notes 128-30.

granted a patent does not get one. At first glance, then, the doctrinal basis of a false negative is irrelevant.

Still, even if the doctrinal basis of a false negative does not affect the kind of problem the mistake causes, it might affect the magnitude of the problem. This could occur if (1) the doctrines produce false negatives at different rates on patents that differ by important characteristics and (2) inventors know about the doctrinal bases of false negatives. At the most general level, if some doctrines produce false negatives on trivial inventions and other doctrines produce false negatives on important inventions, the latter will have more serious consequences for innovation incentives than the former.

To see how systematic variation might occur, note first that false negatives will more likely occur on hard questions than on easy ones. For any given doctrine, there will likely be some patents that are close to the line and some that are clear-cut. Some patents will plainly satisfy the nonobviousness requirement; others will force the patent system to resolve difficult technical and legal issues in order to evaluate whether that requirement is satisfied. And, any given patent may present hard questions on some doctrines and easy questions on others. If the rate at which patents present hard questions on the various doctrines depends on important characteristics of those patents, then the doctrinal basis of the false negative will matter. The next two subsections will consider two characteristics of inventions that could plausibly be the source of systematic variation in the rates at which they pose hard questions on different patentability rules: the value of the invention and the technological field of the invention.

(a) *Value-Based Variation*

We might first wonder whether the patentability rules vary with respect to the value of those patents that present close questions. Suppose there are two kinds of patents. Call one set Edison patents.¹⁰⁸ These are very valuable patents, providing effective coverage over inventions that have very high market value. Call the next set Shark Suit patents.¹⁰⁹ These are patents that, although valid, are simply not worth very much, perhaps because the inventions they cover have little market value. Variation in false negative rates by doctrine across Edison and Shark Suit patents would cause variation in the costs of false negatives by doctrine. Patentability rules that produce many Edison patent false negatives (and few Shark Suit ones) would have serious consequences for incentives to innovate because they increase the likelihood that inventors who make major

108. See U.S. Patent No. 223,898 (filed Nov. 4, 1879).

109. See U.S. Patent No. 4,833,729 (filed Mar. 13, 1985), reproduced in Fromer, *supra* note 96, at 600-06.

advances will not be compensated. Patentability rules that produce many Shark Suit patent false negatives (and few Edison ones) would have negligible consequences for incentives to innovate because they only increase the likelihood that inventors who make trivial contributions will not be compensated.

For example, we might consider the possibility that patents that present close questions on novelty will be systematically less valuable than those that present close questions on nonobviousness. After all, the invention must be new before we can even consider whether it is an obvious variation on what is already known. If it's not even clear that the invention is new, it must be quite similar to the preexisting state of the art and therefore of little additional value.

This pattern is not, however, inevitable. Consider *Forest Labs., Inc. v. Ivax Pharms., Inc.*¹¹⁰ That case involved the antidepressant Lexapro, which earned several billion dollars in revenue over its life¹¹¹—we can think of this as a case involving an Edison patent. The patent covered Lexapro's active ingredient, S-citalopram, which is the mirror image of another compound, R-citalopram.¹¹² These two compounds are naturally found in what is known as a racemic mixture—a 50-50 combination of the two mirror-image structures—and could not easily be separated or produced in purified form.¹¹³ One prior art reference identified the racemic mixture and predicted that R-citalopram would be more potent than S-citalopram.¹¹⁴ The easy question was nonobviousness—no one at the time thought that obtaining S-citalopram (as opposed to R-citalopram) would be desirable, and, if the field did not already possess it by virtue of the reference's identification of the racemate, there was no known way to obtain it.¹¹⁵ The difficult question related to novelty—whether the prior art reference, which disclosed the racemic mixture but did not provide any instruction on how to separate the compounds, meant that the state of the art included each of the purified compounds.¹¹⁶ This case thus helps show that valuable patents can pose hard novelty questions and easy nonobviousness ones.

So while there is superficial appeal to the suggestion that novelty false negatives occur more frequently on low-value inventions and

110. 501 F.3d 1263 (Fed. Cir. 2007).

111. *Forest Labs., Inc. v. Ivax Pharms., Inc.*, 438 F. Supp. 2d 479, 494-95 (D. Del. 2006).

112. *See Forest Labs.*, 501 F.3d at 1265-66. These mirror-image compounds are known as enantiomers. There are different ways of identifying enantiomers that produce different naming conventions. Here, S-citalopram is the same as (+)-citalopram and R-citalopram is the same as (-)-citalopram. *Id.* at 1265.

113. *See id.* at 1265-67.

114. *Id.* at 1267.

115. *See id.* at 1269.

116. *See id.* at 1267-69.

that nonobviousness false negatives occur more frequently on high-value inventions, it is less compelling on further examination. Close questions on novelty generally do not arise because the alleged invention is so similar to the preexisting state of the art that it's unclear whether there are any differences between the two; instead, they arise because it is unclear what was known in the first place.¹¹⁷ Close novelty questions are common when the scope of preexisting knowledge is difficult to ascertain *ex post*.¹¹⁸ Put differently, those working in the field did not realize what they (might have) had, and so it is unclear whether they had it at all. Because they did not realize what they had (or that it mattered whether they had it), the non-obviousness question is easy—no one would have thought to obtain the invention. But because no one realized the import of the invention, it is possible that it was sitting right there all along, and hence a close novelty question arises.

True, *Forest Labs* does not conclusively demonstrate that the patentability rules cannot produce false negatives at different rates for inventions of different values. It may still be the case that some patentability rules are much more likely to produce false negatives on Edison patents than on Shark Suit patents, and that other patentability rules have other tendencies. This is ultimately an empirical question, but it does not appear that variation in the value of inventions that produce false negatives on different doctrines is compelled as a matter of logic or the structure of the patentability rules.¹¹⁹

(b) *Industry-Specific Variation*

Although the doctrines may not generally vary with respect to the value of the inventions for which they produce false negatives, they might vary with respect to the technological field of the inventions for which they produce false negatives.¹²⁰ Utility, for example, is almost never cited as a basis for a rejection outside of the chemical and biological fields, so it may produce more false negatives on chemical and biological patents than other ones.¹²¹ If so, false negatives on utility will more likely reduce incentives to invest in the chemical and biological industries than in other fields because inventors working in those areas have more reason to worry that excessively stringent ap-

117. See Seymore, *supra* note 21, at 931-36 (describing difficulties in ascertaining whether a prior art reference is enabled for purposes of novelty analysis).

118. See *id.*

119. See Meurer, *supra* note 5, at 687-88 (concluding that questions about whether particular doctrines produce mistakes of different magnitudes are “questions for future empirical research”).

120. See generally Burk & Lemley, *supra* note 9 (describing cross-industry variation in the application of the patentability doctrines).

121. *Id.* at 1644-46 (citing “biology and chemistry” as the “only exceptions” to the patent system’s general abandonment of the utility requirement).

plications of the rule will affect their ability to patent their inventions. Similarly, the patent system might be interpreting nonobviousness such that it is difficult for software patents to meet, but easy for biotech patents to meet.¹²² If so, then false negatives on nonobviousness will more likely reduce incentives to invest in the software field than the biotech field.

This selection by industry is not compelled by the nature of the doctrines themselves or the structure of the patentability rules; instead, it is caused by the way in which the courts are applying the patentability rules. We could imagine applying the utility doctrine in an industry-neutral way such that the utility requirement produces the same rate of close questions regardless of the type of invention.¹²³ That's not to say that we necessarily should apply the patentability rules in an industry-neutral way; it is only to say that to the extent that we apply the rules in an industry-specific way, false negatives on different doctrines will have industry-specific consequences. Because the effects of false negatives on different doctrines will be industry-specific, the costs of false negatives on different doctrines will vary.

The impact of a false negative also depends on the efficacy of non-patent appropriation mechanisms available to inventors who are wrongly denied patents.¹²⁴ These mechanisms—including trade secrecy, tacit knowledge, trademarks, and contracts—reduce the costs of false negatives by providing other ways for inventors to appropriate returns to their inventions.¹²⁵ When those mechanisms are about as effective as patents and are not much more costly, false negatives will have little effect on incentives to innovate. When those mechanisms are much less effective than patents or are much more costly, false negatives will have significant effects on incentives to innovate. The efficacy of these mechanisms might vary by industry, thus compounding the industry-specific selection effects of the patentability rules. Because the availability of non-patent appropriation mechanisms also depends on the timing of the false negative, I explore that issue in detail in the next subsection.

122. See generally Dan L. Burk & Mark A. Lemley, *Is Patent Law Technology-Specific?*, 17 BERKELEY TECH. L.J. 1155, 1156 (2002).

123. See Burk & Lemley, *supra* note 9, at 1645-46 (concluding that differences in the application of the utility doctrine “[are] not reflected in the statute but derives ultimately from judicial interpretation”).

124. See generally Jonathan M. Barnett, *Private Protection of Patentable Goods*, 25 CARDOZO L. REV. 1251, 1251-52 (2004).

125. *Id.*

2. *Timing Effects*

The ordinary intuition might be that early false negatives are more costly than late false negatives.¹²⁶ Given the inventor's compliance with all of the patentability rules, we know that the optimal scenario would be to grant her a twenty-year patent term.¹²⁷ Because a late false negative confers a longer period of exclusivity (from patent grant to judicial invalidation of the patent in infringement litigation) than an early false negative, a late false negative comes closer to the optimal scenario. All else equal, we should therefore prefer late false negatives.

This subsection complicates that intuition. First, early false negatives are harder for inventors to observe than late ones. An early false negative is therefore less likely than a late false negative to affect inventors' beliefs about the likelihood that they will be faced with false negatives on future inventions. Second, early false negatives might facilitate inventors' efforts to use non-patent appropriation mechanisms to mitigate the costs of the mistake; late false negatives might frustrate such efforts. As a result, we should be more willing to make early false negative mistakes than the ordinary intuition would suggest.

(a) *Observability of False Negatives*

False negatives are costly because they affect inventors' incentives to innovate. But they can only do so if inventors are aware of them. Generally, it will be easier for inventors to learn of late false negatives than early ones. There are only about one hundred patent cases that make it to trial in a given year, and late false negatives will occur in a subset of those cases.¹²⁸ There are over 450,000 applications filed annually, and early false negatives will be made on a subset of those.¹²⁹ Whatever the actual rate of mistakes in each instance, there will almost surely be many more early false negatives than late ones.

Moreover, inventors will need to know not only that a patent has been invalidated or an application rejected; they will also need to assess the basis for that rejection in order to know whether it was justified. It will be much easier to do so for late false negatives than early

126. Although the doctrinal basis of a false negative will likely affect its costs, the kinds of problems do not vary by doctrine. Accordingly, I refer here generally to false negatives.

127. 35 U.S.C. § 154 (providing for a patent term of twenty years from the date of filing).

128. See Lemley, *supra* note 3, at 1501.

129. See United States Patent and Trademark Office Patent Technology Monitoring Team, *supra* note 97 (reporting that over 450,000 patent applications have been filed a year since 2006 and that in 2010 520,277 applications were filed). Lemley and Sampat estimate that the PTO rejects about 17% of the unique applications it receives. Mark A. Lemley & Bhaven Sampat, Essay, *Is the Patent Office a Rubber Stamp?*, 58 EMORY L.J. 181, 194 (2008).

ones. Litigation resolving patent validity issues will often produce easily-accessible judicial opinions. But for most early false negatives, the only easily-accessible record will be a copy of the application; the interactions between the applicant and the examiner will typically be available only upon request to the PTO.¹³⁰ It's unlikely that inventors will take the initiative to collect and comb through the files of abandoned applications to see whether the patent system has produced an early false negative. Inventors will simply have to do much more to learn of early false negatives than late ones, and it is therefore likely that the impact of any given early false negative will be less than the impact of any given late false negative.

(b) *Availability of Non-Patent Appropriation Mechanisms*

Early false negatives and late ones both destroy inventors' ability to use the optimal appropriation mechanism provided by patent law. But inventors may also use non-patent appropriation mechanisms.¹³¹ For example, an inventor might use trade secrecy to prevent rivals from accessing the information underlying the invention.¹³² The inventor's decision to apply for a patent suggests that the non-patent appropriation mechanisms would be less effective than a patent, but those mechanisms are still better than nothing. As a result, false negatives are somewhat problematic when they force inventors to resort to non-patent appropriation mechanisms; they are more problematic when those mechanisms are unavailable.

There are several non-patent appropriation mechanisms that an inventor might try to use if she were wrongly denied a patent—foremost among them are trade secrecy, tacit knowledge, trademarks, and contracts.¹³³ For each of these, the patent system's evaluations of patentability can be viewed as inputs to the inventor's decision-making process. As the inventor develops a market-ready product based on the invention, she must make decisions about how to design the product, whom to share product information with, how much to invest in branding, and so on. The inventor will make different decisions based on her prediction of whether she will have a patent on her invention. If she knows she will have a patent, she will invest less in non-patent appropriation mechanisms; if she knows she

130. See 37 C.F.R. § 1.14(a)(1)(ii) (2011) (providing access upon request to the "file of an abandoned application that has been published").

131. See generally Barnett, *supra* note 124 (discussing extra-legal protections used by innovators).

132. See generally Mark A. Lemley, *The Surprising Virtues of Treating Trade Secrets as IP Rights*, 61 STAN. L. REV. 311 (2008) (discussing the use of trade secrets to protect intellectual property including technologies that could be covered by patents).

133. This list is not exhaustive; other mechanisms might exist too. But these are the likely alternatives.

will not have a patent, she will invest more. So information about the likelihood of patent protection is a factor in her decision-making process.

Because the non-patent appropriation mechanisms are more effective the sooner they are adopted, the value of the patent system's decisional outputs declines over time.¹³⁴ At the extreme, when the non-patent appropriation mechanisms have become entirely unavailable because, for example, the information can no longer be protected by trade secrecy, the government's informational output is worthless as an input to the inventor's decision-making. The following subsections sketch out the sensitivity of each of these mechanisms to the timing of the false negative and note industry-by-industry variation in the availability of these mechanisms where appropriate.

(i) *Trade Secrecy*

One non-patent appropriation mechanism that an inventor might use is trade secrecy. Trade secrecy and patent protection are imperfect substitutes.¹³⁵ Like patent law, trade secrecy can prevent rivals from acquiring the information needed to copy the invention.¹³⁶ Un-

134. See generally David A. Super, *Against Flexibility*, 96 CORNELL L. REV. 1375 (2011) (arguing that in many contexts, the value of decisional outputs by the government declines over time). An inventor may, of course, adopt at least some of these mechanisms at any time—including between the patent system's initial grant of a patent and its late false negative invalidation of that patent—if she believes there is a sufficiently high risk that the patent system will ultimately produce a false negative. But the patent system's goals are better served if inventors can abandon alternative appropriation mechanisms and rely instead on the patent right to exclude. In many cases where inventors seek patent protection, the alternative mechanisms are costly, second-best tools for solving the public goods problem at the heart of the justification for the patent system. In the extreme case where the patent system commits to only making false negatives early, if at all, inventors could drop alternative appropriation mechanisms as soon as they obtain patents. The arguments in the text illustrate the potential desirability of that extreme case but can also justify the less extreme case in which the patent system commits to a sufficiently high ratio of early false negatives to late false negatives.

135. Patentable information is generally eligible for trade secret protection. See *Kewanee Oil Co. v. Bicron Corp.*, 416 U.S. 470, 491-92 (1974) (holding that Ohio's trade secret law was not preempted by federal patent law because, *inter alia*, "the extension of trade secret protection to clearly patentable inventions does not conflict with the patent policy of disclosure"). The converse is not true—some information that is ineligible for patent protection is nonetheless eligible for trade secret protection. See *id.* at 482-83. Also, though the focus in the text is on the use of the two systems as substitutes, they can also be used as complements, in which some of the information regarding an invention is protected by patent law and some by trade secrecy. I focus here on the information for which patent protection is optimal, as indicated by the inventor's decision to apply for a patent on it. To the extent there is other information that is better protected by trade secrecy, it is outside the scope of this Article.

136. See UNIF. TRADE SECRETS ACT §§ 2, 3 (1985) (providing for injunctive and monetary relief for the misappropriation of trade secrets); see also *Kewanee Oil*, 416 U.S. at 487-88 (noting that inventors with "a legitimate doubt as to [the] patentability" of their inventions may avoid patent law because of the "risk of eventual patent invalidity" and that "[t]rade secret protection would assist those inventors in the more efficient exploitation of their discoveries"); Lemley, *supra* note 132, at 326 (arguing that "[w]e grant rights over secret information for the same reason we grant rights in patent and copyright

like patent law, it cannot protect an inventor against rivals who obtain the information by reverse engineering or independent invention.¹³⁷ The inventor may have to use costlier organizational, manufacturing, or employment strategies to maintain a trade secret than to maintain a patent.¹³⁸ If an inventor has applied for a patent, we know that perhaps because of these differences, patent protection would be the optimal legal regime for her invention, at least by her lights. But even when patent protection is best, trade secrecy may still be a second-best option if the patent system makes a mistake.

Trade secrecy is not equally effective for all inventions. It will be essentially useless when the invention is self-disclosing and therefore easy to reverse engineer.¹³⁹ The classic example here is the paper clip.¹⁴⁰ Everything that the rival needs to know to copy the paper clip is contained in the product, and so the paper clip industry (and other similar industries) might suffer greater harm from false negatives. And even if the invention is not self-disclosing, it may be costly to design the product so as to make reverse engineering difficult.¹⁴¹ Software and consumer electronics are plausible examples of inventions that can be made more or less resistant to reverse engineering

law—to encourage investment in the research and development that produces the information”). These treatments view patent and trade secret protection as alternatives chosen by the inventor *ex ante*; they do not view trade secret law as a potential remedy to mistakes made by the patent system.

137. See RESTATEMENT (THIRD) OF UNFAIR COMPETITION § 43 (1995) (stating that “[i]ndependent discovery and analysis of publicly available products or information are not improper means of acquisition” of a trade secret); Lemley, *supra* note 132, at 319 (“[Anyone] who acquires a trade secret by developing it on her own or by reverse engineering it is free to do what she wants with the secret.”).

138. See, e.g., *Kewanee Oil*, 416 U.S. at 485-86 (describing the measures necessary to protect a trade secret); LANDES & POSNER, *supra* note 9, 329 (arguing that reliance on trade secrets rather than patents “would cause inefficiencies in manufacture”); Jonathan M. Barnett, *Intellectual Property as a Law of Organization*, 84 S. CAL. L. REV. 785 (2011) (arguing that inventors who prefer to outsource production or distribution processes will find it easier to do so if their products are protected by patents); Peter S. Menell, *Bankruptcy Treatment of Intellectual Property Assets: An Economic Analysis*, 22 BERKELEY TECH. L.J. 733, 739-40 (2007) (stating that because the information loses protection once it becomes publicly known, an inventor relying on trade secrecy might “spend an inordinate amount of resources on building high and impervious fences around their research facilities and greatly limiting the number of people with access to the proprietary information” and may have to pay employees more to prevent them from going to competitors).

139. See Lemley, *supra* note 132, at 338-39 (arguing that self-disclosing inventions can be protected by patents, but not by trade secrecy); see also Katherine J. Strandburg, *What Does the Public Get? Experimental Use and the Patent Bargain*, 2004 WIS. L. REV. 81, 104-18 (developing distinction between self-disclosing and non-self-disclosing inventions). For example, the weakness of non-patent appropriation strategies for pharmaceuticals generally has led pharmaceutical firms simply to refuse to develop drugs that seem to have weak or non-existent patent positions. See Benjamin N. Roin, *Unpatentable Drugs and the Standards of Patentability*, 87 TEX. L. REV. 503, 545-56 (2009).

140. Lemley, *supra* note 132, at 338-39.

141. *Id.* at 338-41.

at some cost to the inventor.¹⁴² On the other hand, when the invention is not visible to the world and reverse engineering is very costly, then a false negative will have little effect.¹⁴³ An example here may be chemical process inventions, which may be impossible to discern solely from observing the end result of the process.¹⁴⁴

The timing of the mistake will often determine the inventor's ability to use the second-best option of trade secrecy. Merely applying for a patent does not destroy the availability of trade secret protection, but publication of the application or issued patent does.¹⁴⁵ Applications are typically published eighteen months after filing.¹⁴⁶ The inventor can, however, keep the application secret if she certifies to the PTO that she has not filed and will not file for a foreign patent covering the same invention.¹⁴⁷ If the applicant abandons the application before publication, then the abandoned application remains unpublished, and the possibility of trade secrecy is preserved.¹⁴⁸

A false negative that occurs before publication allows the inventor to use trade secrecy to mitigate the costs of the mistake. Because a late false negative necessarily occurs after a patent has issued, an inventor cannot turn to trade secret law to fix that mistake. Of course, if early false negatives occur after publication, then there is no difference in the viability of trade secrecy as a fix for early and late false negatives—it is simply unavailable for either. But holding all else equal, early mistakes are better here than late ones; inventors can sometimes use trade secrecy to fix the former but can never use trade secrecy to fix the latter.

(ii) *Tacit Knowledge*

Inventors might also try tacit knowledge strategies to deal with false negatives. Tacit knowledge is simply information that has not been written down—instead, it is acquired and transmitted by expe-

142. *Id.*

143. *Id.* at 339-41.

144. *Id.* at 339-40.

145. 4-11 CHISUM, *supra* note 42, § 11.02(4) (“Patent applications, pending or abandoned, may contain trade secrets enforceable under state law.”). Documents detailing the interactions between the applicant and the examiner are also provided upon written request. 37 C.F.R. § 1.14 (2011). Publication destroys the availability of trade secrecy because, upon publication, the information will no longer satisfy the requirement that the trade secret “not be[] generally known.” UNIF. TRADE SECRETS ACT § 1(4) (1985).

146. 35 U.S.C. § 122 (2006); 37 C.F.R. § 1.14.

147. 35 U.S.C. § 122(b)(2)(B).

148. *See* 35 U.S.C. § 122(b)(2)(A)(i) (providing that an “application shall not be published if that application is . . . no longer pending”); 4-11 CHISUM, *supra* note 42, § 11.02(4). This is true unless the inventor cites to or otherwise relies on the abandoned application in an issued patent. 37 C.F.R. § 1.14(a)(1)(iv) (permitting publication of abandoned applications when they “are identified or relied upon”).

rience and observation.¹⁴⁹ Tacit knowledge can be contrasted with codified knowledge, which is knowledge that has been written down or recorded and can be easily transmitted and acquired in that form.¹⁵⁰ The type of information does not determine whether knowledge is tacit or codified; some information will be more costly to codify, but that does not imply that codification is impossible.¹⁵¹ Instead, inventors must choose whether the knowledge they have will be preserved in tacit form or converted into codified form.

For example, suppose an inventor designs a neurological implant that improves a patient's memory. The inventor and her team will likely have a substantial amount of tacit knowledge regarding how to physically implant the device in a patient's brain. If the inventor's rivals cannot physically access her team or observe them implanting the device, it will be difficult for them to acquire the tacit knowledge needed to use the invention well.¹⁵² As this example should make clear, rivals may eventually acquire the knowledge themselves by buying the invention and conducting routine experiments with it. But those experiments will likely be quite costly, especially for a device like a neurological implant. The point of a tacit knowledge strategy is not that it will necessarily prevent rivals from acquiring the information underlying the invention; instead, it is only that it will make it more costly for them to do so.¹⁵³

Tacit knowledge will be more suited to some industries and inventions than others. As a general rule, when the size of the potential market for an invention is large, it will be costlier to adopt tacit knowledge strategies to exclude rivals.¹⁵⁴ Tacit knowledge strategies have low initial costs (because they rely on preexisting stores of knowledge generated during development of the invention), but high marginal costs (because they require the inventor to spend resources

149. See Burk, *supra* note 9, at 1014-16. A simple example of tacit knowledge is a tennis serve. See Robin Cowan & Dominique Foray, *The Economics of Codification and Diffusion of Knowledge*, 6 INDUS. & CORP. CHANGE 595, 606 (1997) (offering the tennis serve example); see also David Foster Wallace, *Federer as Religious Experience*, N.Y. TIMES, Aug. 20, 2006, at A46, available at <http://www.nytimes.com/2006/08/20/sports/playmagazine/20federer.html> (arguing that written language cannot convey an understanding of Roger Federer's game and that only "witnessing, firsthand" will do). Reading a description of how to serve a tennis ball doesn't do much to teach you how to do it; the knowledge can only be acquired by observation and (repeated) experience.

150. Robin Cowan, Paul A. David & Dominique Foray, *The Explicit Economics of Knowledge Codification and Tacitness*, 9 INDUS. & CORP. CHANGE 211, 211-12 (2000).

151. Cowan & Foray, *supra* note 149, at 600.

152. Cowan et al., *supra* note 150, at 222.

153. See Lynn G. Zucker, Michael R. Darby & Jeff S. Armstrong, *Commercializing Knowledge: University Science, Knowledge Capture, and Firm Performance in Biotechnology*, 48 MGMT. SCI. 138, 141 (2002) ("[T]acit knowledge can be viewed as at least partially . . . excludable information and thus 'appropriable' as long as it remains difficult (or impossible) to learn it.").

154. See Cowan et al., *supra* note 150, at 222.

training sales representatives who then must spend time demonstrating to customers how to use the invention).¹⁵⁵ Codified knowledge strategies have the opposite profile: high initial codification costs and low marginal costs of transmission.¹⁵⁶ As this difference suggests, tacit knowledge strategies will work well for industries that produce inventions like medical devices, in which there is typically a lot of interaction between the salesperson and the customer; they will work less well for industries that produce inventions like paper clips, in which there is typically little interaction between the salesperson and the customer.

The tacit knowledge response to a false negative is better implemented sooner rather than later. The inventor of the neurological implant will have to decide the degree to which she will transform her tacit knowledge into codified knowledge by writing user manuals, publications for medical journals, and so on. She might prefer a codified knowledge strategy because it is cheaper if the product will be widely adopted, but she might be worried that the codified knowledge strategy will also lower her rivals' copying costs. Moreover, if the inventor launches the product using largely codified knowledge strategies, she will be unable to return to a tacit knowledge strategy because the preexisting codified knowledge will be freely available to rivals. The inventor will have to make many such choices during the commercialization process. Early false negatives allow inventors to increase reliance on tacit knowledge during commercialization and product launch; late ones may come after the critical decisions are made. So as with the trade secrecy response, tacit knowledge will be a more effective appropriation tool when false negatives are made early rather than late.

(iii) Trademarks

If the information underlying the invention is not susceptible to either trade secret or tacit knowledge strategies, another possible approach is to turn to trademark protection. On the conventional understanding, trademarks lower consumer search costs by letting them rely on experience or recommendations for information about product attributes.¹⁵⁷ Once a consumer has experience with the inventor's trademarked product, she may have to pay some positive

155. *See id.*

156. *See id.*

157. LANDES & POSNER, *supra* note 9, at 174 (stating that trademarks convey information about the source of a product, which "economizes on search costs by lowering the costs of selecting goods on the basis of past experience or the recommendation of other consumers"). This function is especially important when a product has important attributes that are difficult to evaluate at the point of purchase—durability, medicinal efficacy, and taste are some typical examples.

switching costs to try a competitor's version.¹⁵⁸ If so, then the inventor can charge that customer a price equal to the sum of the competitor's price and the customer's switching costs.¹⁵⁹

The inventor's ability to impose switching costs on her customers depends on the absence of viable substitute products; if viable substitutes are available, consumers will simply avoid the inventor's products. This strategy thus depends on the inventor being the first to market. During the time between the inventor's product launch and the launch of the first competing product, the inventor has some period during which customers can experience her products but no competing ones. In that interim, the inventor can, if she anticipates competition, turn to the trademark strategy to impose switching costs on her customers and associate her products with her brand.

The importance of switching costs in purchasing decisions will vary; they will be most important when the trademarked products are either of very high or very low value.¹⁶⁰ For low-value goods, it will usually not be worth it for the consumer to spend time trying to learn about competing products.¹⁶¹ For high-value goods, the risks of trying an alternative product may be large.¹⁶² In either of these scenarios, inventors can use trademarks to preserve barriers to entry.

This strategy will work best when the inventor can anticipate the launch of competing products. The inventor can use the pre-competitive period to establish her brand—during this time, any consumers who buy the product will buy her brand. The period immediately preceding a competitor's product launch will see the inventor spending resources to broaden her customer base, be it through lower prices or increased advertising.¹⁶³ Then, when competitors appear, the inventor will have the largest possible base of customers for whom switching costs are high.

The relationship between the trademark strategy and the timing of false negatives is thus subtle. Unlike trade secrecy and tacit knowledge, the ideal timing of the trademark strategy is related not to the inventor's commercialization process and product launch, but

158. Gideon Parchomovsky & Peter Siegelman, *Towards an Integrated Theory of Intellectual Property*, 88 VA. L. REV. 1455, 1477-78 (2002). Switching costs are affected by several conditions, including whether the customer prefers variety and whether the switching costs will be amortized over many purchases. *See id.* at 1481-84 (listing some factors that affect brand loyalty).

159. *Id.* at 1478.

160. *See* Barnett, *supra* note 124, at 1260-61.

161. *Id.*

162. *Id.*

163. *See* Parchomovsky & Siegelman, *supra* note 158, at 1514-15 (arguing that patentees will try to expand customer bases as patent expiration approaches so that they can use their trademarks to capture consumers with high switching costs in the post-expiration period); *see also id.* at 1489-93 (describing case studies of patentees engaging in this strategy).

instead to her ability to anticipate her competitors' product launches. At first glance, there does not seem to be any reason to suspect that the timing of a false negative will affect the inventor's ability to anticipate competing products.

But recall the setting of these mistakes. Early false negatives occur during prosecution of the patent, when the inventor alone is interacting with the patent system. Assume that the early false negative occurs before the inventor launches her product. If so, then she will know when she launches the product that competitors will enter the market in as little time as it takes them to imitate. In contrast, late false negatives occur during infringement litigation. That means that at least one competing product has been on the market for some period of time.¹⁶⁴ And that competing product launched during a period of time for which the inventor expected exclusivity—after all, the patent was valid and in force. Unlike an early false negative, then, a late false negative comes after the key moment for implementing the trademark strategy—the launch of a competing product. Because of this, it's plausible that trademarks will be better able to fix early false negatives than late ones.

(iv) *Contracts*

In certain instances, an inventor might also be able to form contractual relationships that limit her rivals' ability to offer cheap copies of the invention. If the invention's functionality depends in significant part on some important input, then the inventor may be able to secure large portions of that input before rivals enter the market.¹⁶⁵ Similarly, if some distributors or resellers have an important share of the end user market, the inventor may again seek exclusive relationships that limit market entry.¹⁶⁶ Of course, inventions and industries

164. I set aside the possibility that the inventor will sue an alleged infringer before the infringer starts selling the invention due to infringing activities that occur during product development. Of course, such cases occur, but they are a small part of the overall picture given the difficulty for the patentee in detecting such infringement.

165. A recent example here appears to be Apple's strategy for its iPhone and iPad products. Those products require special glass for their multi-touch functionality, and reports suggest that Apple has formed exclusive relationships with suppliers that have increased barriers to entry. See Robert X. Cringely, *Apple's Money* (Aug. 1, 2011, 7:30 PM), <http://www.cringely.com/2011/08/apples-money/> (speculating that Apple is using its cash reserves to buy "flash RAM and iPhone displays in amounts that move whole markets and guarantee Apple the lowest prices anywhere" and "the most reliable supply," such that "Apple has an effective consumption-side monopoly for certain mobile components."); Matthew Humphries, *Apple Secures 60% of World's Touch Panel Output* (Feb. 17, 2011, 8:30 AM), <http://www.geek.com/articles/gadgets/apple-secures-60-of-worlds-touch-panel-output-20110217/> (describing Apple's purchases of glass used for touch screens and the difficulty that tablet computer rivals have had securing cost-competitive sources).

166. See Barnett, *supra* note 124, at 1263 (describing how "a first-mover may cultivate arrangements with resellers and other retail agents that may include specially tailored and unusually favorable contractual provisions . . .").

will vary in the degree to which important supplies or distribution channels are susceptible to exclusive contractual relationships.

Like the trademark strategy, effective use of contractual strategies depends on the inventor's ability to predict entry by her rivals. If competition is unlikely because a patent has been granted, then the inventor need not incur the potential expense of these contractual strategies. Once entry is on the horizon, though, these exclusive arrangements may be justified. As with trademarks, then, the inventor can more easily use contracts to respond to an early false negative (which provides notice of potential entry) than a late one.

3. *Summary and Caveats*

The doctrinal basis of a false negative can affect its costs by selecting for patents that have importantly different characteristics. The most plausible characteristic is the technological field of the invention. Because the courts apply the patentability rules in an industry-specific manner, the costs of false negatives on some doctrines will fall especially heavily on certain industries.

The ordinary intuition is that, for any given patent, the appropriability problem will be less severe for late false negatives than early ones. This is because in the case of a late false negative, the inventor will have had some period of exclusivity; in the case of an early false negative, she will have had none. The inventor's ability to fix the appropriability problem through non-patent mechanisms complicates that intuition. Trade secrecy, tacit knowledge, trademarks, and contracts can be used to exclude rivals from the information or retain some market power. But because those strategies are easier to implement for early false negatives than late ones, we should be less concerned about early false negatives than the ordinary intuition would suggest. Moreover, because early false negatives are less visible than late ones, they would seemingly have less serious consequences for inventors' expectations about the likelihood of future mistakes.

I have focused on the appropriability problem the inventor faces following a false negative and the tools the inventor might use when the patent system makes mistakes. But in addition to solving that problem, the patent system also aims to promote dissemination of technical information about the invention.¹⁶⁷ Disclosure is more important in cumulative industries, in which each invention builds on many other inventions, than discrete industries, in which each invention essentially stands alone.¹⁶⁸ The non-patent appropriation strate-

167. See *supra* text accompanying notes 25-28.

168. Robert P. Merges & Richard R. Nelson, *On the Complex Economics of Patent Scope*, 90 COLUM. L. REV. 839, 880-84 (1990) (describing, among others, the

gies differ in the degree to which solving the appropriation problem exacerbates the dissemination problem. Successful use of trade secrecy and tacit knowledge limit dissemination; trademarks and contractual strategies do not affect the ability of rivals to learn about the invention. We might therefore be more worried about the use of trade secrecy and tacit knowledge strategies in cumulative industries than in discrete ones. In those situations, the plausible advantages of early false negatives will be diminished.

IV. IMPLICATIONS

The analysis thus far has described how the type, timing, and doctrinal basis of patent mistakes affect their costs. This Part spells out some implications of that analysis. I will first compare the relative costs of patent mistakes. I will then describe applications of this analysis for patent examination priorities and for the long-standing debate regarding the deference that courts owe to PTO decisions.

A. *The Relative Costs of Patent Mistakes*

As previously discussed, it is unclear whether the costs of false negatives on any given doctrine will be higher for early or late mistakes. The answer to this question depends on (1) whether the inventor's ability to use non-patent appropriation tools in response to an early false negative allows her to keep more of the returns to her invention than the short term of exclusivity she would enjoy with a late false negative and (2) whether the more easily-observable nature of late false negatives outweighs any appropriation advantage they enjoy over early false negatives. For present purposes, though, let us assume that we are in an industry in which the non-patent appropriation tools are very effective for early false negatives, and the feedback effects of late false negatives are large.¹⁶⁹

Begin with the disclosure and definiteness rules. For these, I have suggested some reasons to think that late false positives do not add much to the costs of early false positives. Combined with the possibility that early false negatives are less costly than late ones, this indicates that the patent system's late assessment of compliance with the disclosure and definiteness rules should be less stringent than its early assessment—it should be more biased in favor of false negatives early rather than late. Of course, this does not mean that it

pharmaceutical, consumer packing, and toy industries as following a discrete innovation model and the aircraft and semiconductor industries as following a cumulative innovation model).

169. Industry-by-industry variation along these dimensions suggests a possible role for industry-by-industry variation in how the patentability rules are applied. *Cf.* Burk & Lemley, *supra* note 9, at 1675-95 (arguing that courts do and should tailor the application of patent law on an industry-by-industry basis).

should be biased in favor of early false negatives over early false positives. False negatives might always be more costly than false positives, whether made early or late. The point here is simply that the relative assessment changes over time. Because the costs of false negatives increase as we move from the early assessment to the late one—and the costs of false positives decrease—even if we prefer false negatives overall, we should be less optimistic about them when made late rather than early, as compared to false positives.

I have also suggested that for the scope and invention rules, we cannot make any general statements about the costs of false positives over time. Like disclosure and definiteness, the costs of false negatives on scope and invention should be higher for late mistakes than early ones. But the costs of false positives on these rules may also increase from the early assessment to the late one and may even increase more than do the costs of false negatives. As a result, it is unclear how the relative comparison between early false positives and false negatives on scope and invention changes when we move to late false positives and false negatives. Accordingly, while the patent system should be more willing to commit early false negatives on disclosure and definiteness than late ones, there is no clear reason to suspect that such an approach is appropriate for scope and invention rules.

B. Resource Allocation

To make this more concrete, we might consider the following proposal: the disclosure and definiteness rules should only be enforced by the PTO, and defendants should not be able to argue in infringement litigation that a patent is invalid for failure to comply with them.¹⁷⁰ Because most of the costs of false positives on disclosure are incurred early in the patent's life, and because *Markman* hearings resolve the important definitional ambiguities that make definiteness false positives costly, strict enforcement of these doctrines at the late stage does little to reduce ongoing costs while increasing the risk of late false negatives. Assuming the costs of late false negatives are high because they are highly-visible and make resort to non-patent appropriation mechanisms difficult, there is little to be gained from that increased risk.¹⁷¹ It therefore seems plausible that we should allo-

170. The recently enacted America Invents Act implements a limited version of this proposal by preventing a defendant in an infringement case from using the best mode doctrine as a potential basis for an invalidity defense. See Leahy-Smith America Invents Act, Pub. L. No. 112-29, § 15, 125 Stat. 284, 328 (2011) (providing that “the failure to disclose the best mode shall not be a basis on which any claim of a patent may be canceled or held invalid or otherwise unenforceable”).

171. Though a complete assessment of the costs of mistake avoidance is beyond the scope of this Article, we might suspect that the PTO is well-situated to evaluate at least the enablement and definiteness rules during examination. While the invention rules impose serious informational burdens on the PTO to identify relevant prior art,

cate relatively more resources to enforcing these rules at the PTO than in the courts. And, at the same time, more resources could be spent in litigation assessing compliance with the scope and invention rules.

To be sure, there are countervailing considerations. The PTO will inevitably make some mistakes on the disclosure and definiteness rules, and it may be unfair to hold defendants liable for those mistakes; the inventor is almost surely the lowest-cost avoider, so it makes sense to ensure that she has good incentives to avoid them. Moreover, the adversarial nature of litigation might be well-suited to correcting PTO mistakes on these doctrines. The proposal does illustrate, however, a concrete way to apply the results of the analysis here.

C. *The Presumption of Validity*

In a similar vein, consider the Supreme Court's recent decision in *Microsoft Corp. v. i4i Ltd. Partnership*.¹⁷² The Court held that an alleged infringer must demonstrate that a patent is invalid by clear and convincing evidence.¹⁷³ It also rejected the possibility that the lower preponderance standard would apply if the defendant presented evidence that was not available to the PTO.¹⁷⁴ But the Court nonetheless allowed that "the challenger's burden to persuade the jury of its invalidity defense by clear and convincing evidence may be easier to sustain" with new evidence than with evidence that was before the PTO.¹⁷⁵

Implicit in this suggestion is a doctrine-specific approach to the presumption of validity. New evidence will often take the form of references that show the state of the art was more advanced than the PTO had thought; the patent is therefore more likely to have violated the invention rules. But it would be a rare case in which new evidence of invalidity affects the assessment of whether the patent complies with the disclosure and definiteness rules because that assessment is largely conducted within the four corners of the patent document. As a result, the evidentiary burden the defendant must overcome to demonstrate invalidity on the disclosure and definiteness rules will never be "easier to sustain," but his burden to demonstrate invalidity on the invention rules might be.

The Court's approach might be partially justified by the difference between the invention rules on the one hand and the disclosure and definiteness rules on the other. As described here, for the disclosure and definiteness rules, we should be more tolerant of false positives than false negatives at the late stage compared to our tolerance at

enablement and definiteness are self-contained inquiries, requiring only that the examiner understand what's written in the patent document itself.

172. 131 S. Ct. 2238 (2011).

173. *Id.* at 2242, 2244.

174. *Id.* at 2244.

175. *Id.* at 2251.

the early stage; we lack, however, a good reason to take the same approach for the invention rules. Lowering the presumption of validity for the invention rules, but not for the disclosure and definiteness rules, implements essentially this idea.

V. CONCLUSION

This Article has assessed how three important characteristics affect the costs of patent mistakes. Those characteristics—type, timing, and doctrinal basis—make some mistakes more worrisome than others. Because empirical evidence on the costs of patent mistakes is difficult to acquire, the design of the patent system must flow from theoretical arguments. Of course, the arguments here are incomplete—I have not, for example, resolved the debate about the role of disclosure in the patent system. Nor have I said much about the costs of mistake-avoidance; a complete analysis of patent mistakes would include not only the costs of the mistakes, but also the costs of avoiding them. Still, I have evaluated one side of the ledger and spelled out some of the intuitions that would lead us to favor some mistakes over others. The arguments presented here, which are based on areas of widespread agreement within patent theory, thus suggest a way forward through an inherently uncertain and hotly contested area.