Rethinking Prosecution History Estoppel

Douglas Lichtman†

Under the rule of prosecution history estoppel, patent applicants who amend their claims during the course of patent prosecution assume a significant risk: the risk that a court will later construe the changes as concessions that should be read to limit patent scope. This risk is exacerbated by strong evidentiary presumptions under which courts are to assume, unless the patentee presents sufficient evidence to the contrary, that every change triggers estoppel, and that the resulting estoppel forfeits everything except that which the revised claim language literally describes. The justification for these presumptions is that, implemented in this fashion, prosecution history estoppel makes patent scope more predictable. In this Essay, I argue that the benefit comes at too high a price. Drawing on a large empirical study of patent prosecution, I show that, because of these evidentiary presumptions, estoppel is dangerously sensitive to differences among patent examiners and differences across technology categories. That is, estoppel treats similar applications in dissimilar ways, not because of differences on the merits, but instead because of the personal characteristics of the examiners involved and because of differences inherent to the types of technology at issue. A better rule, I argue, would minimize the significance of examiner and technology disparities by reversing the current evidentiary presumptions and thus recognizing estoppel only where there is clear evidence that the applicant and the examiner intended to forfeit a given scope of coverage.

INTRODUCTION

Patent prosecution is an iterative process, and during that process applicants often change the language of their proposed claims. A running debate in patent law considers whether and how evidence of those language changes should be used in litigation. On one view, the meaning of a word can be distorted when taken out of context, and the best way to put patent language into context is to study the history of the patent document. On another view, evidence drawn from a patent's prosecution history is cumbersome, ambiguous, sometimes mis-

---

† Professor of Law, The University of Chicago. Sincere thanks to workshop participants at the American Bar Foundation, George Washington University, Harvard, Michigan, and the University of Chicago. For helpful comments on earlier drafts, I also thank Marty Adelman, Lisa Bastarache, Will Baude, Erica Benton, Paul Janicke, Bill Landes, Mike Meurer, Dick Posner, Eric Posner, Max Schanzenbach, Lior Strahilevitz, Cass Sunstein, John Thomas, and Polk Wagner. Dennis Crouch provided invaluable assistance in gathering the empirical data presented here. Lastly, special thanks to Judge Randal Rader, who not only made much appreciated contributions to this Essay while participating in the George Washington University workshop, but also has since taken up the torch in his own *Festo IX* concurrence. See *Festo Corp v Shoketsu Kinzoku Kogyo Kabushiki Co* (*Festo IX*), 344 F3d 1359, 1374–77 (Fed Cir 2003) (Rader concurring). Comments welcome at dgl@uchicago.edu.

leading, and often incomplete, and the goals of the patent system would therefore be better served were courts to ignore language changes and focus instead on the final claim language standing alone. The puzzle bears an obvious resemblance to a perhaps more familiar question in statutory interpretation; namely, whether and how legislative history should be used to construe the language of an enacted statute.

In the patent context, the debate has primarily played out in the shadow of the doctrine of equivalents. The doctrine of equivalents empowers courts to construe patent claims to cover not only that which they literally describe, but also some range of equivalent subject matter that technically falls outside the literal claim language but on policy grounds seems appropriately considered part of the patent holder's exclusive domain. The doctrine is typically invoked in instances where unscrupulous competitors would otherwise be able to undermine the patent grant by exploiting loopholes in the literal claim language.[1] Loopholes eligible for this sort of protection include loopholes caused by the unavoidable imprecision of language,[2] loopholes caused by events and circumstances that were not reasonably foreseeable at the time the literal claim language was drafted,[3] and loopholes where the accused invention is an insubstantial variant of the invention literally described.[4] As these examples make plain, a major drawback to equivalents analysis is that it renders uncertain the precise boundaries of any particular patent claim. One mechanism used to address that worry—and the most controversial means through which the history of the patent document influences claim interpretation[5]—is the rule of prosecution history estoppel.[6]

---

1 As the Supreme Court explained in *Graver Tank & Manufacturing Co v Linde Air Products Co*, 339 US 605 (1950), the "essence of the doctrine is that one may not practice a fraud on a patent" by making "unimportant and insubstantial changes" that, "though adding nothing, would be enough to take the copied matter outside" the scope of the literal claims. Id at 607–08.

2 See, for example, id at 607 (suggesting that without equivalents, patentees would be "at the mercy of verbalism"); *Festo Corp v Shoketsu Kinzoku Kogyo Kabushiki Co* (*Festo VIII*), 535 US 722, 734 (2002) ("[T]he doctrine of equivalents is premised on language's inability to capture the essence of innovation.").

3 See, for example, *Pennwalt Corp v Durand-Wayland, Inc*, 833 F2d 931, 938 (Fed Cir 1987) (en banc) ("[T]he facts here do not involve later-developed computer technology which should be deemed within the scope of the claims to avoid the pirating of an invention."), revd in part on other grounds, *Cardinal Chemical Co v Morton International, Inc*, 508 US 83 (1993).

4 See, for example, *Carman Industries, Inc v Wahl*, 724 F2d 932, 942 (Fed Cir 1983) (stating that equivalents analysis is appropriate where the accused infringer seeks "to appropriate the invention with minor modification to avoid the literal language of the claims").

5 The history of the document is also used to clarify the literal meaning of patent claim language. This use, however, is relatively uncontroversial. See *Zodiac Pool Care, Inc v Hoffinger Industries, Inc*, 206 F3d 1408, 1414 (Fed Cir 2000) (explaining the difference between intrinsic and extrinsic evidence).

6 There are other doctrines in patent law explicitly designed to reduce the uncertainty

Prosecution history estoppel applies when an applicant during patent prosecution narrows a claim "to avoid the prior art, or otherwise to address a specific concern . . . that arguably would have rendered the claimed subject matter unpatentable."[7] In these instances, estoppel bars the applicant from later invoking the doctrine of equivalents to recapture the lost ground. As the Supreme Court put it in *Schriber-Schroth*[8]—and note how the Court's explanation sounds in classic estoppel and waiver terms—an applicant "may not, by resort to the doctrine of equivalents, give to an allowed claim [the] scope which it might have had without the [narrowing] amendments."[9] By amending the claim, the applicant is deemed to have "recognized and emphasized the difference between the two phrases and proclaimed his abandonment of all that is embraced in that difference."[10]

For estoppel to achieve its purpose of reducing the uncertainty inherent in equivalents analysis,[11] estoppel itself must be implemented in a predictable fashion. Courts have therefore built into the rule heavy evidentiary presumptions. For example, although as a technical matter prosecution history estoppel applies only where a narrowing amendment was made to satisfy a requirement of the Patent Act—and note how broad a category that already is—the Supreme Court held in *Warner-Jenkinson*[12] that the patent holder bears the burden of establishing the reason for any narrowing amendment, and, where no explanation can be established, courts are to presume that estoppel applies.[13] This has proven to be a difficult presumption for patent holders

---

created by the doctrine of equivalents. For discussion, see Martin J. Adelman, et al, *Cases and Materials on Patent Law* 798–841 (West 2d ed 2003).

7 *Warner-Jenkinson Co v Hilton Davis Chemical Co*, 520 US 17, 30–31 (1997). See also *Festo VIII*, 535 US at 735–36. While all the major cases focus on claim language amendments, an applicant can trigger estoppel in other ways. For example, the act of deleting a claim of broader scope than those ultimately allowed carries with it implicit representations that may later be held against the applicant. Indeed, mere arguments may also give rise to estoppel, even if unaccompanied by any language changes. See Adelman, et al, *Patent Law* at 818 (cited in note 6) (noting the "established principle" that "prosecution history estoppel does not require an amendment").

8 *Schriber-Schroth Co v Cleveland Trust Co*, 311 US 211 (1940).

9 Id at 221.

10 *Exhibit Supply Co v Ace Patents Corp*, 315 US 126, 136 (1942).

11 Although this is the theory on which the courts routinely focus, estoppel does arguably serve other purposes. See notes 66–67 and the text accompanying and immediately preceding.

12 *Warner-Jenkinson Co v Hilton Davis Chemical Co*, 520 US 17 (1997).

13 See id at 33. It is unclear what presumption, if any, is applied in answering the threshold question of whether a given amendment narrows or broadens a claim. That is, there are two preliminary questions to ask with respect to estoppel: (1) whether the change narrowed or expanded claim scope, and (2) if the amendment is a narrowing one, whether the narrowing was done to satisfy a requirement of the Patent Act. The Supreme Court has yet to rule on the proper presumption to apply in the first of these two inquiries, and the Federal Circuit apparently views that as an open question as well. See *Festo IX*, 344 F3d at 1366–67 (discussing the effects of the *Warner-Jenkinson* and *Festo VIII* presumptions on the second question, but remaining silent as to the appropriate basis for determining whether an amendment should be construed as

to overcome given that, historically, neither patent examiners nor patent applicants have put much effort into documenting the reasons for patent claim amendments.[14] Moreover, the presumption sticks even if it turns out that the amendment at issue was not in fact necessary to preserve patent validity. For instance, if the examiner incorrectly interprets the prior art and, because of that error, the applicant agrees to narrow a given claim, the applicant is still bound by the concession.[15] Thus, in practice, the only evidence that immunizes a patentee is clear evidence that a given narrowing amendment was not made with the *intent* to preserve claim validity; evidence that the change was not *necessary* to satisfy Patent Act requirements is not enough.[16]

Similarly, when estoppel applies, courts must determine its scope; here again, courts employ a strong evidentiary presumption. Specifically, under the Supreme Court's *Festo* decision,[17] the patentee bears "the burden of showing that the amendment does not surrender the particular equivalent in question."[18] The patentee can carry this burden by showing that the equivalent was unforeseeable at the time of the claim amendment,[19] or that the "rationale underlying the amendment . . . [bore] no more than a tangential relation to the equivalent in question."[20] But where the applicant cannot make these or similar showings—and, again, this is likely given how poorly patent prosecution is documented under current Patent Office practices[21]—the doctrine of equivalents is in essence repealed and the applicant must rely on literal claim coverage alone.

In this Essay, I raise two concerns over the practical implications of the modern estoppel rule. The first is based on an empirical finding

---

narrowing).

14 For a rich discussion of current practice, see John R. Thomas, *On Preparatory Texts and Proprietary Technologies: The Place of Prosecution Histories in Patent Claim Interpretation*, 47 UCLA L Rev 183, 188–91, 203 (1999).

15 See *Exhibit Supply*, 315 US at 137. Of course, the applicant can resist the examiner and ultimately appeal the dispute to higher officials at the Patent Office. See note 50.

16 The evidence also must come from documents on file at the Patent Office, rather than from documents in the patentee's private possession. See *Festo IX*, 344 F3d at 1369. Any other rule would make it difficult for rivals to determine patent scope prior to litigation, thus undermining predictability.

17 *Festo VIII*, 535 US at 722.

18 Id at 740. The Federal Circuit previously proposed an even stronger presumption. See *Festo Corp v Shoketsu Kinzoku Kogyo Kabushiki Co* (*Festo VI*), 234 F3d 558, 564 (Fed Cir 2000) (holding, prior to Supreme Court reversal, that when a claim amendment creates estoppel, "no range of equivalents is available for the amended claim element"), vacd and remd, 535 US 722 (2002).

19 *Festo VIII*, 535 US at 740.

20 Id.

21 See note 14. See also *Festo VI*, 234 F3d at 575 (describing context-sensitive estoppel as "unworkable" and thus advocating a bright-line rule). Of course, all this might change now that estoppel is such a harsh rule, although such a change would be costly given the large number of applications filed with the Patent Office each year.

that there are statistically significant differences between patent examiners in terms of their tendency to require that patent applicants alter claim language. That is, while some examiners routinely insist on significant language alterations, others regularly leave the original claim language largely intact. These differences are of substantial magnitude and they persist even after controlling for factors such as the type of technology involved and the length of the original application. They are troubling because they cause the burdens of prosecution history estoppel to be distributed in an arbitrary fashion. If an applicant happens to be assigned an examiner who tends to require few language alterations, estoppel is not much of a risk. If an applicant happens to be assigned an examiner who tends to demand a large number of language alterations, by contrast, the threat of estoppel looms large. From the perspective of the patent applicant, this difference is a random factor, unrelated to the merits, that forces unlucky applicants either to suffer the harms associated with estoppel or to spend additional resources during patent prosecution resisting the examiner and documenting with care any amendments ultimately made. From a policy perspective, meanwhile, these examiner disparities mean that using prosecution history estoppel to reduce the uncertainty associated with equivalents analysis has a large and unanticipated cost: it makes the patent system more random, specifically by linking patent scope to what turns out to be a personal characteristic that varies considerably from one patent examiner to another.[22]

My second concern derives from a related empirical finding that there are also statistically significant differences across technology categories in terms of the extent to which claim language is changed during the average patent prosecution. Claim language describing a patentable advance in nanotechnology, for example, is on average altered much more significantly than is claim language describing a patentable advance related to automobile engines or electrical lighting. It is not entirely clear what drives these differences. But if it is right to assume that language adjustments are more common in complicated and rapidly evolving technologies—technologies about which it is more difficult for applicants to write appropriate claims in the first instance, and technologies where there is more room for reasonable disagreement between applicant and examiner at the time of patent prosecution—then estoppel threatens the doctrine of equivalents in the very cases where equivalents analysis is needed most: instances where conscientious applicants working with qualified examiners

22 Note that examiners rarely directly propose claim language. Instead, an examiner influences claim language indirectly by refusing to accept a literal claim until the claim is worded in what the examiner deems to be an acceptable manner.

might still fail to capture in literal terms the proper boundaries of the invention at hand.

These concerns can be addressed. Were the aforementioned evidentiary presumptions reversed, for example, prosecution history estoppel would be triggered less often, and, at that, only when the applicant and examiner actually meant to foreclose a given equivalent. This would reduce the legal risk associated with amendments to claim language, and it would therefore mitigate the impact of differences among examiners and across technology categories. Another approach would have the Patent Office take more seriously its role in documenting the process of patent prosecution. This might be expensive, but it, too, would help ensure that estoppel would be triggered only when the applicant and examiner actually meant to foreclose a given equivalent, again rendering estoppel less sensitive to examiner and technology disparities. At the same time, new mechanisms could be introduced to fulfill estoppel's intended role in terms of reducing the uncertainty created by the doctrine of equivalents.

Naturally, there is much more to say on all these points. I proceed as follows. In Part I, I introduce the basic methodology behind my empirical work, explaining the data set and identifying strengths and weaknesses in my approach. In Part II, I present my core statistical analysis. I show that the identity of the examiner drives the extent of claim language alteration, and that claim language alterations also differ significantly from technology to technology. Finally, in Part III, I discuss the implications of these technology and examiner disparities, developing in further detail the challenges they raise for the rule of prosecution history estoppel and, through that rule, for the doctrine of equivalents as well.

I. METHODOLOGY

In November 2000, the United States Patent and Trademark Office initiated a program under which newly filed patent applications are made public "after the expiration of a period of 18 months from the earliest filing date for which a benefit is sought."[23] Not all applications are published under this new program. For instance, the Patent Office does not publish an application if either the applicant abandons the application during that eighteen-month period[24] or the applicant asserts a special exemption that maintains confidentiality for patent applications that have been filed in the United States but have not been filed in a foreign country that itself requires disclosure after

[23] 35 USC § 122(b) (2000). The program operates under the authority of the American Inventors Protection Act of 1999, Pub L No 106-113, 113 Stat 1501A-552.

[24] 35 USC § 122(b)(2)(A)(i).

eighteen months.[25] Nevertheless, in under three years, the program has already generated a public archive of more than 400,000 patent applications, and new applications are being added at the rate of approximately 20,000 per month.

In January 2003, I collected the 300,000 applications then available and traced each through its time at the Patent Office.[26] The goal was to identify applications for which I could take a single issued patent, compare that patent to the single application at hand, and in that way detect any changes that were made during the process of patent review. This approach saves the expense and labor associated with gathering such information directly from the records kept at the Patent Office, making practicable an empirical study that otherwise would have been cost-prohibitive.[27] To identify the appropriate applications, I obviously needed to exclude applications still under review. I also needed to exclude applications that either had splintered into multiple related patents, or were still eligible to do so, because in those instances the proper interpretation of any application/patent pair would have been ambiguous.[28] This filtering process left me with almost 20,000 workable applications.[29]

---

25 Id. The applicant must also certify that the invention will not be subject to such an application in the future. Id.

26 The Patent Office website provides all the necessary information, particularly through the Patent Application Information Retrieval (PAIR) system. See http://pair-direct.uspto.gov (visited Dec 16, 2003).

27 See 37 CFR § 1.19 (2002) (listing the fee schedule for obtaining patent prosecution paperwork).

28 A patent application can split into multiple related applications either by the filing of a divisional application, which literally draws material first included in the original application and divides it into two or more separate applications, or by the filing of a continuation-in-part, which takes material from the original application but in addition introduces new material. Both variants complicate the otherwise intuitive comparison approach, and ultimately I decided that it was better to exclude such applications rather than introduce error by attempting to incorporate them.

Note that I also excluded continuation applications. A continuation application is an application submitted some time after patent prosecution has begun. It typically revises its associated original application and is separately identified primarily as a way of collecting higher fees from applicants who make large numbers of changes. I excluded continuations because the baseline of interest is the application as it was first submitted, not the application as it appeared after some interaction with the relevant examiner. Even where I excluded a continuation application, however, I included the original application that led to that continuation, so long as the original application otherwise qualified.

29 Some small number of these applications possibly should have been excluded because, as late as one month before publication, the Patent Office allows an applicant to swap his original application for an updated version. See 37 CFR § 1.215(c) (2002). Swapped applications are not distinguished from other applications when published; the only way to detect them is to consult the relevant file wrappers. They are problematic for this project because the baseline of interest is the application as it was first submitted, not the application as it appeared after some interaction with the relevant examiner. That said, unpublished Patent Office statistics suggest that fewer than 100 applicants take advantage of this regulation each year, which means that, at most, about 5 percent of our applications were affected. See email from Stephen Kunin, Deputy Com-

Of course, building the data set in this manner introduced some biases. For example, the data set does not include any patent applications for which patent prosecution took fewer than eighteen months, because patent applications are published under the new regulations only after eighteen months have passed. The data set likewise excludes patent applications abandoned by the applicant or denied by the examiner, because the comparison strategy works only for applications that actually led to issued patents. The data set also excludes any application that qualified for the special exemption mentioned above, and any application that was splintered either voluntarily or at the direction of a patent examiner. None of these exclusions is particularly troubling for current purposes, however, because each primarily biases the data *against* what turned out to be my primary findings. That is, each of these exclusions reduces diversity in the data set, making it all the more surprising that I found statistically significant differences among examiners and across technology categories.

A. Examiner Disparities

To answer the question about examiner disparities, I needed to draw from each application/patent pair three basic types of information. The first type of information was simply the name of the examiner or examiners who actually reviewed each application in the data set. This was easy information to acquire, given that examiners are identified by name on the patents they allow. Indeed, I only had to decide how to code patents for which two examiners—an inexperienced "assistant examiner" working under a more experienced "primary examiner"—together reviewed the application. In those situations, I ultimately decided to treat each unique team as a separate entity, reasoning that, from the perspective of a patent applicant, an evaluation conducted by examiners Smith and Jones is meaningfully different from an evaluation conducted by examiners Smith and Williams, even though both pairs include examiner Smith. In the database, I therefore gave each unique team, and each individual examiner, a distinct identifier.

The second type of information that I needed to extract was some measure of the extent to which a given application's claim language changed during the course of patent prosecution. To that end, I decided to count (1) the number of unique words used in the issued claims but not in the original claims, and (2) the number of unique words used in the original claims but not in the issued claims. More precisely, I made a list of the vocabulary used to describe the inven-

missioner for Patent Examination Policy (Jan 25, 2003) (on file with author).

tion in the application claims, and I made a list of the vocabulary used to describe the invention in the patent claims; I then compared the two lists, counting any word that was present on one list but missing from the other. The intuition is that every time an applicant either introduces or removes a vocabulary word during patent prosecution, he assumes the risk that a court will later construe the change as a concession. The number of vocabulary changes is therefore a rough proxy for estoppel risk.[30]

The third type of information I needed to gather was information that might help to control for relevant differences among the applications. One obviously important control was some measure of application length. I ultimately decided to use for this purpose the number of different vocabulary words present in the original application claims. This is one measure of length, and it correlates strongly to other obvious measures of length such as the number of words in the application claims, the number of words in the application overall, and the number of claims listed in the application. This count has an added virtue, however, in that it also provides information about the complexity of the original application; applications with a high degree of vocabulary diversity are likely harder to evaluate than applications in which the number of distinct vocabulary words is relatively low.[31]

---

30 Admittedly, there are limitations to this approach. For example, it accidentally counts typographical errors that are present in the application but corrected in the issued patent, even though obvious error corrections entail little estoppel risk. It also counts language changes where the applicant has other (broader) claims that cover the same subject matter and, hence, there is a plausible argument that the changes should not be read to forfeit any ground. Moreover, this approach necessarily fails to detect any estoppel not associated with language changes, such as an estoppel that arises by virtue of an argument presented by the applicant during patent prosecution. These and related limitations are admittedly important, but their implications should not be overstated. After all, these factors affect every observation, and thus they are unlikely to distort comparisons among examiners and across technology categories significantly.

That said, I have run a number of robustness checks to look for these sorts of problems. For example, in one run I coded not only the word counts referenced above, but also weighted versions where the introduction or removal of a rare word counted more heavily than the introduction or removal of a common one. Specifically, from a sample of 10,000 issued patents, I created a frequency table that showed the number of patents in which any given word appeared. I then assigned scores based on the inverse of the frequencies, such that the loss or addition of a common word like "the" or "said" was scored close to zero, whereas the loss or addition of a rare word like "hand-activated" or "vacant" was scored close to one. I ended up dropping these weighted vocabulary counts from the analysis, however, because they turned out to be almost perfectly correlated with the simpler unweighted tallies. Other robustness checks—for example, a run that counted only those language changes that affected independent claims—similarly seemed to have little effect on the ultimate results.

31 Another reason for using the number of vocabulary words rather than other intuitive measures is that the other measures are each significantly distorted by the pyramid structure of patent claiming. Applicants draft broad "independent" claims to stake out the main invention, and then largely redundant "dependent" claims to repeat the theme of each broad claim and add additional narrowing information. The number of claims therefore constitutes a crude measure of the length of an application; dependent claims inflate the total even though they are typically

A second obviously important control was information regarding the type of technology described in each application. If an application claiming an advance in nanotechnology naturally invites more language alteration than an otherwise comparable application relating to automobile bumpers—an outcome I confirm in this study—comparing examiners without simultaneously accounting for technology introduces significant error into the analysis: an examiner who works more often on automobile bumpers would seem less exacting than his nanotechnology peer, no matter what the real differences between the two. The difficult question was how best to capture this information. After all, every successful application to some degree describes its own distinct technology, and yet it is impossible to conduct statistical analysis where every observation is considered a unique group.

In the patent literature, approaches vary, with some papers introducing elaborate classification schemes that distinguish hundreds of technology categories, while others settle for relatively coarse alternatives that lump all technologies together under six or ten headings. I decided to err on the side of caution and adopted one of the more fine-grained approaches. Specifically, the Patent Office classifies issued patents according to a system that distinguishes 421 technology classes; I borrowed this system, using the ten classes for which I had the most observations to study examiners one technology at a time. The ten classes that I used are listed in Table 1.

trivial to evaluate once their related independent claims have been studied. Counting the number of words used in the application claims also poorly measures length, again because that number is unduly sensitive to the number of dependent claims. While counting the number of independent claims seems a better option, that tally ignores the fact that dependent claims do add new information that must be reviewed, albeit less information than would be added by an additional independent claim. Thus, the number of distinct vocabulary words seemed like the most reliable measure of application length, in that it is not overly sensitive to repetition and yet does account for any new vocabulary introduced in dependent claims.

TABLE 1
The ten patent classes studied in this research; descriptions are drawn from the Patent Office classification chart

Patent Class	Description
438	Semiconductor Device Manufacturing Process
365	Static Information Storage and Retrieval
257	Active Solid State Devices
439	Electrical Connectors
123	Internal Combustion Engines
327	Miscellaneous Active Electrical Nonlinear Devices, Circuits & Systems
359	Optics: Systems and Elements
361	Electricity: Electrical Systems & Devices
347	Incremental Printing of Symbolic Information
701	Data Processing: Vehicles, Navigation, and Relative Location

I initially suspected that, in addition to controlling for application length and technology, it would also be important to control for the size and expertise of the law firm, if any, that prosecuted each application. This might be important if, for example, the patent prosecution strategies adopted by large firms differ from those adopted by patent boutiques in ways that affect claim language alteration. Issued patents typically identify the law firm that represented the applicant during prosecution, so I recorded law firm names whenever they were available, and then matched them to publicly available information about the number of patents each firm prosecuted in the last five years, the approximate number of licensed patent attorneys employed by each firm, and the average experience level of the patent attorneys employed by each firm.[32] I was able to gather this information for approximately two-thirds of the applications in the database; yet, to my surprise, the results lacked explanatory power. While these variables were statistically significant in the context of an occasional patent class, even there the effect was always several orders of magnitude smaller than the effects attributed to the various examiner-specific variables. I therefore decided to simplify my regressions by dropping these controls.

32 I counted the number of patents prosecuted by each firm myself, and I acquired the information about the number of licensed attorneys and their experience from PatentRatings, LLC, which in turn was able to get the information from records maintained and made public by the Patent Office through its Office of Enrollment and Discipline.

More broadly, when I designed this project, I identified a large number of controls that I imagined as possibly relevant. For example, it likely matters whether the person who actually drafted the patent application met with the patent examiner in person, rather than interacting exclusively through telephone calls and the exchange of written documents. The number of prior art sources cited in each application would also likely be informative, because a long list might signal that the applicant was particularly diligent in preparing the application, or that the application falls into a relatively crowded art. It might also be valuable to know the country where the claimed technology was first developed or patented, because an application written originally for another country's patent system might differ substantially from one originally drafted with Patent Office rules and regulations in mind. It might even be helpful to know if there was a company involved in guiding the application, as applications prosecuted on behalf of individual inventors surely differ from applications for which a for-profit corporation is paying the fees and calling the shots.

Rather than overwhelm the analysis with an unending list of considerations, however, I decided instead to restrict the study to include only those examiners for whom I had ten or more observations. The logic is that, within a given technology class, most other factors are randomly distributed across applications, such that, over the course of a large enough sample, every examiner working within a particular technology class will face approximately the same number of applications originally drafted for a foreign country, approximately the same number of applications for which the applicant requests in-person negotiations, and so on. If true, these factors can be safely ignored, as they will not distort comparisons from one examiner to another.[33]

---

[33] Of course, I would be even more comfortable with this assumption were I able to set the minimum threshold at twenty or thirty observations, rather than ten. However, two factors cautioned against such an adjustment. First, the higher the threshold, the more data excluded from the analysis. That is a serious cost in this study given that I started out with only approximately 20,000 patent applications representing nearly 400 patent classes and nearly 3,000 patent examiners. Second, excluding examiners with few observations biases the data against a finding of examiner diversity. The reason is that a minimum threshold excludes examiners who work slowly and thus could not process the requisite number of applications during the time frame under consideration, and also excludes examiners who work quickly and thus processed more than the requisite number but did not have enough that lasted the eighteen months required to trigger mandatory publication. The higher the threshold, the greater these distortions. Thus, I was reluctant to choose too high a threshold, especially given that most of the missing controls are likely of trivial import as compared to the technology, length, and complexity factors that I explicitly account for in the regressions.

TABLE 2
The data, sorted by patent class and counting only those examiners with ten or more observations

Patent Class	Examiners	Observations
438	11	157
365	17	451
257	11	140
439	12	250
123	16	334
327	15	244
359	12	187
361	8	144
347	10	173
701	10	202

Table 2 describes the resulting data. The first column identifies the relevant technology class, the second column reports the number of examiners working in that class who processed ten or more applications during the timeframe of interest, and the third column counts the total number of applications processed by those examiners.

B. Technology Disparities

The preceding information not only permits the study of differences among examiners, but also permits the study of differences across technologies. Indeed, the only adjustment I made to the data to facilitate the latter inquiry was to replace the 421-category technology classification system developed by the Patent Office with a 36-category alternative developed by the National Bureau of Economic Research.[34] The rationale for the change is purely cosmetic: the 36-category approach uses classifications that might be more intuitive for the lay reader. The categories, and the number of observations per category, appear in Table 3.

34 Online at http://www.nber.org/patents (visited Dec 16, 2003).

TABLE 3

The data set by NBER classification

Category	Subcategory	Observations
Chemical	Agriculture, Food, Textiles	58
Chemical	Coating	218
Chemical	Gas	87
Chemical	Organic	364
Chemical	Resins	324
Chemical	Miscellaneous	1,115
Computers/Comm	Communications	911
Computers/Comm	Computer Hardware/Software	842
Computers/Comm	Computer Peripherals	411
Computers/Comm	Information Storage	787
Drugs/Medical	Drugs	302
Drugs/Medical	Surgery/Medical Instruments	292
Drugs/Medical	Biotechnology	112
Drugs/Medical	Miscellaneous	101
Electrical/Electronic	Electrical Devices	1,258
Electrical/Electronic	Electrical Lighting	325
Electrical/Electronic	Measuring/Testing	507
Electrical/Electronic	Nuclear/X-rays	183
Electrical/Electronic	Power Systems	1,163
Electrical/Electronic	Semiconductor Devices	1,217
Electrical/Electronic	Miscellaneous	354
Mechanical	Materials Processing	471
Mechanical	Metal Working	329
Mechanical	Motors, Engines, Parts	979
Mechanical	Optics	713
Mechanical	Transportation	749
Mechanical	Miscellaneous	692
Other	Agriculture, Husbandry, Food	179
Other	Amusement Devices	119
Other	Apparel/Textile	298
Other	Earth Working/Wells	130
Other	Furniture/House Fixtures	329
Other	Heating	151
Other	Pipes/Joints	134
Other	Receptacles	186
Other	Miscellaneous	1,164

II. ANALYSIS

I report my statistical methods and findings in three parts. The first part articulates a simple regression model and uses it to test whether examiner identity influences patterns of claim language alteration. The second part uses that model to estimate the magnitude of any examiner-specific effects. Lastly, the third part introduces a comparable model that tests whether and to what degree technology affects these same language considerations.

A. Examiners Matter

I define an examiner's editorial "STYLE" to be the examiner's proclivity to alter patent vocabulary, expressed as a fraction in which the numerator is the sum of the number of new vocabulary words introduced in the patent claims plus the number of existing vocabulary words omitted from the original application claims, and the denominator is the total number of vocabulary words used in the original application claims. Style is thus a percentage measure of vocabulary change; a larger score implies more significant language alterations.

My regression model can be specified as follows:

$$\text{STYLE}_{\text{Application}} = \text{STYLE}_{\text{Examiner}} + \bullet$$

where $\text{STYLE}_{\text{Application}}$ is the editorial style reflected in the application/patent pair at issue, $\text{STYLE}_{\text{Examiner}}$ is the idiosyncratic editorial style of the relevant examiner, and • stands in for error and unobserved inputs. Technology is not referenced in the equation because, as explained in the previous Part, I control for technology in this part of the study by focusing on one patent class at a time.

I used the median test to determine whether examiner identity influences the style variable.[35] As those familiar with this sort of statistical work know, the median test is not a powerful test; it often fails to detect patterns even when they are in fact present. The test therefore is not useful for ruling out the possibility of a pattern, but it is particularly useful for establishing the existence of a pattern. The median test has another virtue as well: it makes few assumptions about the distribution of the data under consideration. ANOVA, by contrast, is widely used in the literature, but it is reliable only where the groups being tested are all drawn from populations that have the same approximate variance. The Kruskal-Wallis test is another common choice, but it is inaccurate when applied to data for which a large number of the observations take on the same value.

---

35 For background on the median test, see Jack Johnston and John DiNardo, *Econometric Methods* 444–45 (McGraw-Hill 4th ed 1997).

Table 4 reports the results of the median test for each of the ten technology classes I considered. The numbers represent the confidence level for the hypothesis that grouping by examiner is not the same as grouping randomly.

TABLE 4

Median test results by patent class

Patent Class	Examiners	Observations	Median Test
438	11	157	0.197
365	17	451	***0.000***
257	11	140	***0.025***
439	12	250	***0.006***
123	16	334	***0.000***
327	15	244	***0.000***
359	12	187	***0.013***
361	8	144	0.541
347	10	173	***0.008***
701	10	202	***0.013***

In summary, even with a relatively insensitive test, the data regarding language alterations suggest that examiner identity matters in five of the ten technology categories at a confidence level greater than 0.01, and in eight of the ten categories at a confidence level greater than 0.05. This is compelling evidence that there are examiner-specific effects.[36]

B. Magnitude Estimates

To estimate the magnitude of the various examiner effects is admittedly difficult using my data, both because there is a great deal of noise in the patent process, and because publication of patent applications is such a new program that at this stage I have a very limited number of observations per examiner. That said, I report here the point estimates derived from running tobit regressions for each of my ten technology classes.[37] In each regression, the style of the relevant observation was the dependent variable, and the independent vari-

---

36 This evidence is consistent with the results obtained by other researchers who have looked to see whether examiners vary along dimensions other than their tendency to alter claim language. See, for example, Iain M. Cockburn, Samuel Kortum, and Scott Stern, *Are All Patent Examiners Equal? Examiners, Patent Characteristics, and Litigation Outcomes*, in Wesley M. Cohen and Stephen A. Merrill, eds, *Patents in the Knowledge-Based Economy* 19 (National Academies 2003) (arguing, among other things, that some examiners are more likely than others to have their patents invalidated by the Federal Circuit).

37 For background on the tobit regression, see G.S. Maddala, *Limited-Dependent and Qualitative Variables in Econometrics* 149–96 (Cambridge 1983).

ables were dummy variables standing in for the style of each examiner working in the relevant patent class. I used a tobit regression because, in a somewhat surprising 20 percent of the observations, no changes were made to claim language during patent prosecution and thus the data were bunched at zero.[38] Again, the point estimates are significantly imprecise, yet they nevertheless communicate some information about the magnitude of each examiner effect, and, perhaps more importantly, they suggest that magnitude differences across examiners are not trivial and might indeed be quite sizeable.

Table 5 summarizes the results. For each technology class, I include the mean style for applications in that class; the difference between the examiner in the relevant sample who edits most and the examiner in the relevant sample who edits least, reported as a percentage of the mean; and the difference between the examiner who is at the 75th percentile and the examiner who is at the 25th percentile, again reported as a percentage of the mean.

TABLE 5

Magnitudes of the various examiner effects

Patent Class	Mean	High–Low	75th–25th
438	0.13	70%	34%
365	0.09	185%	70%
257	0.20	190%	50%
439	0.25	154%	35%
123	0.10	333%	159%
327	0.21	166%	60%
359	0.16	216%	66%
361	0.21	134%	16%
347	0.15	359%	100%
701	0.11	299%	76%

38 Tobit is the correct choice because observations are censored at zero, not truncated there. Intuitively, the tobit regression in this context recognizes that, while two examiners might both allow a given application to go through unchanged, the examiners might nevertheless differ substantially in terms of how close a call that was. The tobit regression thus acknowledges that differences among examiners can be obscured by seemingly equivalent scores of zero.

The results reveal considerable differences among examiners. Even the conservative 75th–25th measure suggests that, on average, 66 percent of the style score is determined solely by the identity of the examiner involved.[39]

C. Technology Matters

To study whether technology influences the number of language alterations made to a given application, I repeated the above analysis but grouped applications by technology rather than examiner identity. More precisely, I used the median test to ask whether sorting the data by technology produced a pattern of results inconsistent with random grouping, and I ran tobit regressions using dummy variables that represented not the examiners, but the technology types. As mentioned earlier, for this part of the study and for purely cosmetic reasons, I report the results using the more intuitive 36 technology categories suggested by the National Bureau of Economic Research instead of the 421 categories developed by the Patent Office.

---

39 One concern with this statistic is that there might be an informal norm at the Patent Office under which the least complicated applications are assigned to inexperienced examiners and the most complicated applications are reserved for their more experienced peers. This norm would be difficult to maintain, however. At first blush, it is not so easy to predict which applications will prove difficult and which straightforward. Moreover, even if such distinctions can be drawn, complicated applications must be assigned to junior examiners when all the relevant senior examiners are already swamped with work, or when only the junior examiner possesses the appropriate technical expertise. Nevertheless, if inexperienced examiners are assigned a disproportionate share of the straightforward applications, the statistic reported above would be misleading, because differences among examiners would, under that assumption, not necessarily be evidence of examiner-specific variation, but might instead simply reflect the fact that different examiners work on applications of different complexity.

To address this worry, I repeated the median test and regressions reported thus far, but did so using a data set that excludes any examiners working in teams. As I pointed out before, inexperienced examiners do not work alone. Instead, for the first five or six years of employment, an examiner must consult a more senior colleague before marking a patent application as ready for allowance. By excluding issued patents signed by two examiners, I therefore excluded all inexperienced examiners from the study and removed any taint that might be due to seniority-based application allocation. The results: whereas evaluation of the original data regarding language alterations suggested that examiner identity matters in five of the ten technology categories at a confidence level greater than 0.01, evaluation of the data set that excludes inexperienced examiners suggests that examiner identity matters in six of the ten technology categories at a confidence level greater than 0.01. The point estimates were also comparable, with the average style discrepancy rising from the 66 percent figure reported above to 75 percent using data from only experienced examiners.

TABLE 6
Point estimates by technology category

Category	Subcategory	Style
Chemical	Agriculture, Food, Textiles	0.357
Chemical	Coating	0.292
Chemical	Gas	0.205
Chemical	Organic	0.265
Chemical	Resins	0.273
Chemical	Miscellaneous	0.224
Computers/Comm	Communications	0.219
Computers/Comm	Computer Hardware/Software	0.248
Computers/Comm	Computer Peripherals	0.222
Computers/Comm	Information Storage	0.103
Drugs/Medical	Drugs	0.465
Drugs/Medical	Surgery/Medical Instruments	0.167
Drugs/Medical	Biotechnology	0.486
Drugs/Medical	Miscellaneous	0.256
Electrical/Electronic	Electrical Devices	0.188
Electrical/Electronic	Electrical Lighting	0.120
Electrical/Electronic	Measuring/Testing	0.202
Electrical/Electronic	Nuclear/X-rays	0.108
Electrical/Electronic	Power Systems	0.131
Electrical/Electronic	Semiconductor Devices	0.190
Electrical/Electronic	Miscellaneous	0.206
Mechanical	Materials Processing	0.194
Mechanical	Metal Working	0.218
Mechanical	Motors, Engines, Parts	0.070
Mechanical	Optics	0.126
Mechanical	Transportation	0.223
Mechanical	Miscellaneous	0.182
Other	Agriculture, Husbandry, Food	0.225
Other	Amusement Devices	0.300
Other	Apparel/Textile	0.212
Other	Earth Working/Wells	0.224
Other	Furniture/House Fixtures	0.234
Other	Heating	0.230
Other	Pipes/Joints	0.152
Other	Receptacles	0.255
Other	Miscellaneous	0.184

The median test confirms that technology matters at a confidence level exceeding 0.000. Table 6 shows the associated point estimates. The difference between the most- and least-edited technologies is nearly double the average score for technologies taken as a whole, and the difference between the technologies at the 75th and 25th percentiles is approximately 30 percent of the average style score.[40] Note that these point estimates are much more reliable than the point estimates reported with respect to examiners, because this time each is derived using a large number of observations.

III. Discussion

The empirical evidence presented in the previous Part documents two basic insights: (1) that patent examiners differ sharply in terms of their tendency to instigate claim language alterations; and (2) that patterns of claim language alteration also vary substantially from one technology to another. As outlined in the Introduction, these findings have important implications for the rule of prosecution history estoppel. The finding with respect to patent examiners suggests that the risks associated with prosecution history estoppel are allocated arbitrarily; the threat of estoppel grows with the number of claim language amendments, but the number of claim language amendments turns largely on the editorial tendencies of the examiner. Meanwhile, the finding with respect to technology suggests that, while prosecution history estoppel is framed as if it were a rule that applies uniformly across technology categories, it is in fact a rule with technology-specific implications.[41] It creates minefields in industries where language changes are common, but is all but irrelevant for industries where language changes are relatively rare. In this Part, I develop these concerns and integrate them into a broader discussion of both prosecution history estoppel and the doctrine of equivalents.

A. Examiner Disparities

It is hardly surprising that patent examiners vary in ways that affect the scope and value of the patents they issue. Patent examiners are human, after all, and thus like judges, jurors, voters, and English teachers, their personalities and capabilities inevitably affect the decisions they make while on the job. What is surprising is that the rule of prosecution history estoppel is implemented in a way guaranteed to exacerbate the problem. Bluntly, the evidentiary presumptions currently in place render estoppel, and hence patent scope, remarkably sensitive to the happenstance of examiner identity. This is an unantici-

---

40 The difference between the most- and least-edited technologies is 0.417, while the mean for the entire data set is 0.228. The difference between the 25th and 75th percentiles is 0.064.

41 This supports Professors Dan Burk and Mark Lemley in their recent argument that the patent system has many such uneven doctrines. See Dan L. Burk and Mark A. Lemley, *Policy Levers in Patent Law*, 89 Va L Rev 1575 (2003).

pated cost associated with the modern estoppel rule; my basic argument here is that this cost must be weighed against whatever benefits the rule—and specifically the evidentiary presumptions—otherwise provides.

Economic-minded readers might initially reject my analysis on the ground that, so long as the patent system offers the optimal level of protection on average, and so long as patent applicants are in general risk-neutral, examiner inconsistency is not a problem because it will not alter applicant behavior. That is, if an applicant has a 50 percent chance of being awarded a patent that is too broad, and a 50 percent chance of being awarded a patent that is equally too narrow, a risk-neutral applicant will invest in patent-eligible research at the exact level he would under a system in which every patent came out just right. Consistency, on this argument, is irrelevant to patent system design, and thus examiner inconsistency in the context of prosecution history estoppel might seem equally irrelevant, at least at first blush.[42]

One obvious response is that many patent applicants are risk averse. Technology startups, for instance, are surely constrained by the practical and financial concerns associated with unpredictable patent rights and, indeed, unpredictable potential patent liabilities. Moreover, the United States patent system is explicitly designed with the small inventor in mind. The wisdom of this emphasis is subject to challenge, but the descriptive reality is that many patent doctrines unique to the American system—most notably, the rules that award patent protection to the first inventor to conceive of an invention, rather than the first inventor to file for patent protection[43]—intentionally favor small inventors. In the United States patent system, then, small inventors play a substantial role, and risk aversion has a seat at the policy table.[44]

Risk aversion is not the only reason why consistency is in fact an important objective in patent system design. First, even if the possibility of an overly broad patent perfectly offsets the possibility of an overly narrow patent from the perspective of a would-be patentee, it

---

42 The discussion above brackets distributional issues as well as incentives to engage in add-on research. Both of those are obviously very sensitive to patent scope and, hence, both support my argument that examiner consistency does matter from a public policy perspective. For a richer introduction to the economics of add-on innovation, see Robert P. Merges and Richard R. Nelson, *On the Complex Economics of Patent Scope*, 90 Colum L Rev 839 (1990); Douglas Lichtman, *Property Rights in Emerging Platform Technologies*, 29 J Legal Stud 615 (2000).

43 See 35 USC §§ 102(g), 135.

44 Of course, research is itself significantly uncertain, and thus risk-averse patent applicants have other reasons to avoid the patent system beyond the legal uncertainty considered here. See F.M. Scherer, *The Innovation Lottery*, in Rochelle Cooper Dreyfuss, Diane Leenheer Zimmerman, and Harry First, eds, *Expanding the Boundaries of Intellectual Property: Innovation Policy for the Knowledge Society* 3, 11 (Oxford 2001) (documenting the phenomenon associated with patent law in the United States whereby "[a] minority of 'spectacular winners' appropriate the lion's share of total rewards").

does not necessarily follow that the social costs also offset. Quite the opposite, the social costs associated with overly broad patents likely overwhelm the social benefits associated with unduly narrow ones. The details depend on exactly what it means for a patent to be broad versus narrow; but the intuition follows from the familiar principle that, at prices near the monopolistic level, a marginal increase in price imposes more social harm than it yields in patentee benefit,[45] whereas, at lower prices, the ratio of patentee benefit to social harm is typically more favorable if not reversed. Put another way, under a variety of conditions the increase in deadweight loss associated with raising a patent holder's profits by $10 is larger than the reduction in deadweight loss associated with decreasing that patent holder's return by the same $10. Thus, variance that leaves patent applicants indifferent might nevertheless be unattractive from a social welfare perspective.[46]

Along a similar theme, while it is easy to hypothesize a system in which variance is increased but patent applicants on expectation earn the optimal reward, in practice such a system is almost impossible to design. The reason is edge conditions. An applicant who under the optimal system would have been denied patent protection will, under a high variance system, sometimes be awarded protection. But there is no offset against which to cancel that distorted incentive, because there is no such thing as negative patent protection. The reverse will be true for applicants who, under the optimal system, would have received the broadest possible patent for their invention. Here, there is again nothing to offset an errant patent, this time because by definition there can be no patent broader than the patent to which the applicant was already entitled. As a practical matter, then, a system with variance cannot perfectly mirror the outcomes achieved by a more consistent regime. Instead, variance inevitably means increased investment in inventions that would be excluded from protection under

---

45 For discussion and a formal model, see generally Ian Ayres and Paul Klemperer, *Limiting Patentees' Market Power without Reducing Innovation Incentives: The Perverse Benefits of Uncertainty and Non-Injunctive Remedies*, 97 Mich L Rev 985 (1999).

46 Consider a specific example. Define demand to be linear demand of the form $p=-q+1$, where p is price and q is quantity. Suppose that marginal cost is zero and that the optimal patent would give the patent holder sufficient market power such that his price would be 0.3 and thus his producer surplus would be 0.21. If an overbroad patent lets the patent holder charge 0.45 and thus earn a producer surplus that is 0.0375 greater, the corresponding overly narrow patent would allow the patent holder to charge 0.2216 and thereby earn 0.0375 less. By design, then, the patent holder would be indifferent between a patent regime that consistently allowed him to charge 0.3, and a patent regime that half the time allowed him to charge 0.45 and half the time allowed him to charge 0.2216. But society is not indifferent. The latter approach leads to an expected deadweight loss that is 40% greater than the deadweight loss associated with the optimal patent. The numbers: at a price of 0.2216, the producer earns 0.1725 and imposes deadweight loss of 0.0246; at a price of 0.3, the producer earns 0.21 and imposes a loss of 0.045; and at a price of 0.45, the producer earns 0.2475 and imposes a loss of 0.10125.

ideal circumstances and decreased investment in inventions that would be prized most heavily under the optimal regime.

I have focused thus far on difficulties inherent in the assumption of symmetric error, but another concern is that applicants will alter their behavior in socially undesirable ways even if examiner idiosyncrasies do cancel out. For instance, applicants are more likely to delay discretionary investments associated with their inventions in a system with high variance than they are in a more consistent alternative. The reason is that, in a regime with high variance, delay yields information that might in turn help the applicant better prioritize different possible investments. Concretely, an applicant who is unsure whether his claim will cover all touch-sensitive computer screens or merely touch-sensitive computer screens built using a particular design cannot know how best to allocate his marketing and manufacturing resources. Such an applicant will have an incentive to hold those resources in reserve until uncertainty is reduced, in that way increasing the odds that any additional investments will maximize the value of the patent as it ultimately issues.

Another undesirable behavior change is that applicants who would have chosen to pursue patent protection under a consistent patent regime might, in light of examiner variance, opt instead to rely on trade secrecy. Trade secrecy is a substitute for patent protection,[47] and the ideal patent system would be tailored to ensure that appropriate inventions are directed toward the appropriate system. Examiner variance distorts the optimal allocation by changing the patent system's risk-reward profile. The patent system thus must either lure marginal inventions back by compensating in some other way—presumably at the cost of some other social interest—or accept the fact that increased variance distorts the allocation of inventions between these two regimes. Note that there is an even more troubling possibility lurking here: it is possible that applicants can enter the patent system, begin to interact with their assigned examiner, and then retreat to trade secrecy in those instances where the examiner appears stingy. This cherry-picking undermines any argument that examiner inconsistencies cancel out, as generous examiners on this theory end up issuing many more patents than do their more finicky peers.[48]

---

47 See William M. Landes and Richard A. Posner, *The Economic Structure of Intellectual Property Law* 326–33, 354–71 (Belknap 2003) (emphasizing that patent law is in many ways a response to the economic problems associated with trade secret protection).

48 A similar problem is raised to the extent patent applicants file the same or similar applications multiple times, choosing to proceed only when the assigned examiner seems sufficiently generous.

On a related theme, note that, in some arts, there are so few examiners qualified to evaluate patent applications that the examiner's identity is predictable. In those instances, prosecution history estoppel does not increase uncertainty, but it does affect applicant behavior in unintended

The list of problems potentially raised by examiner disparities can go on at some length. Inconsistency might undermine confidence in the patent system as both policymakers and the public realize that patent scope turns significantly on the luck of examiner assignment. It might undermine the statutory presumption of patent validity,[49] this time because courts might become less willing to defer to what they perceive as, at best, a noisy information stream. It might also lead to increased litigation, if, for example, overly broad patents are more likely to be challenged than are patents of appropriate scope. These many reasons combine to suggest that examiner inconsistencies—and patent doctrines that magnify the legal significance of those inconsistencies—are a problem not only for patent applicants, but also for the patent system more generally.

The discussion above disregards the fact that, in the context of prosecution history estoppel especially, applicants can mitigate the consequences of variance by investing additional resources in patent prosecution. For instance, an applicant can appeal adverse examiner decisions up the Patent Office hierarchy, and in certain instances can turn for relief to the federal courts.[50] An applicant can even ask the Patent Office to reopen an issued patent and reconsider the language of its claims.[51] Moreover, as a last resort, applicants can always at a minimum carefully document the reasons for any language changes, in that way creating evidence that can later be used to rebut the various evidentiary presumptions that give estoppel its principal bite. From the applicant's perspective, however, none of this changes the basic point: whether it is because the costs of patent prosecution increase, or because the risk of estoppel grows more severe, it is still true that, the more finicky the assigned examiner, the lower the returns to the patent holder.

From a policy perspective, the analysis also remains largely unchanged even factoring in these various applicant responses. For example, in the discussion above, I point out that, even where the possi-

---

ways. For instance, if the examiner most likely to evaluate patents relating to a particular photographic process is known to require significant claim language alterations, applicants might shy away from patenting inventions in that category, preferring instead to rely on trade secrecy. In most instances, however, it is difficult to predict the identity of the examiner, because the Patent Office employs a large number of examiners, and because there is substantial examiner turnover.

49 See 35 USC § 282 (obligating courts to presume that issued patents are valid).

50 For an introduction to the various options, see Robert Patrick Merges and John Fitzgerald Duffy, *Patent Law and Policy: Cases and Materials* 1153–1254 (LexisNexis 3d ed 2002). It is important not to overstate this point, however, as examiners do retain considerable discretion both because the appellate system is only so fine-grained, and because some decisions cannot be appealed at all. See Thomas, 47 UCLA L Rev at 204–06 (cited in note 14).

51 Although the options here are limited. See Martin J. Adelman and Gary L. Francione, *The Doctrine of Equivalents in Patent Law: Questions That* Pennwalt *Did Not Answer*, 137 U Pa L Rev 673, 716–17 (1989) (discussing restrictions currently in place and advocating reform).

bility of an overly broad patent perfectly offsets the possibility of an overly narrow patent from the perspective of a would-be patentee, the social costs do not necessarily offset. The reason is that the increase in deadweight loss associated with overly broad patents often exceeds the decrease in deadweight loss associated with unduly narrow ones. This same basic logic holds if, instead of receiving unduly narrow patents, unlucky applicants spend more money on patent prosecution but then end up with patents of appropriate scope. After all, under this assumption, the patent system must still generate some number of overly broad patents to compensate for the risk of expensive prosecution. But this time there is no possibility of an offsetting social gain, both because no unduly narrow patents issue, and because the extra money invested by unlucky applicants is itself deadweight loss—which is to say that these resources could be conserved if only examiners were more consistent.

Take stock of what all this means. Because examiners differ considerably in terms of their tendency to require amendments to patent claim language, and because every amendment to patent claim language carries with it some risk of ultimately triggering prosecution history estoppel, the happenstance of examiner assignment has serious implications for patent scope. Draw a finicky examiner, and not only might that examiner directly press for literal claims that narrowly describe the invention at hand, but, by virtue of estoppel, that examiner might also indirectly constrain the protection that otherwise would be available under the doctrine of equivalents. This is obviously troubling to the unlucky patent applicant who is assigned a finicky examiner. My point in this Part is that it is troubling from a public policy perspective as well. Because applicants will adjust their behavior in light of this random effect, they will in many instances choose patterns of investment, disclosure, and prosecution that reduce social welfare as compared to the patterns that would obtain were estoppel not an issue.

B. Technology Disparities

Patterns of claim language alteration might vary from one technology to another for many plausible reasons. It might be, for example, that different industries use patents in different ways, and that those differences cause applicants in some fields to edit patent language more aggressively.[52] It might be that more money is spent by applicants in certain industries than is spent by applicants in other industries, a

---

52 One example: in a field where patents are primarily used defensively, the precise claim language might matter less than it would in a field where patent protection is more directly valued.

difference that again would likely be reflected in the prosecution strategies played by the respective applicants.[53] It might even be that a broad claim is worth more in certain industries than it is in others, a difference that might make the relevant applicants more willing to file overly broad claims even if overly broad claims increase the risk of narrowing amendments and hence estoppel. If part of the explanation, however, is that language adjustments are more common in complicated or rapidly evolving technologies—technologies about which it is more difficult for applicants to write appropriate claims in the first instance, and technologies where, at the time of patent prosecution, there is more room for reasonable disagreement between applicant and examiner—then the implication is that estoppel threatens to repeal the doctrine of equivalents in the very cases where that doctrine is needed most.

To see this, consider the primary policy rationales that support the use of equivalents analysis. The first, and the one that courts most often stress, is the idea that patent holders should in certain situations be protected from "unscrupulous copyists" who would otherwise undermine the value of patent protection by exploiting literal loopholes in patent claim language. The classic articulation comes from the Supreme Court in *Graver Tank & Manufacturing Co v Linde Air Products Co*,[54] where the majority opined that the "essence of the doctrine is that one may not practice a fraud on a patent" by making "unimportant and insubstantial changes" that, "though adding nothing, would be enough to take the copied matter outside" the scope of the literal claims.[55] Many factors determine when this rationale applies; the Court vaguely stated in *Graver Tank* that "equivalency must be determined against the context of the patent, the prior art, and the particular circumstances of the case" and is not "the prisoner of a formula."[56] But a review of the cases suggests that a finding of infringement by equivalents is more attractive (1) the more that loopholes of the type under consideration would otherwise substantially reduce patent value in the long run; (2) the more costly it would be for applicants to anticipate and avoid such loopholes in the future; and (3) the more competitors had adequate notice that the patent would be interpreted to cover the equivalent at issue.[57]

---

53 The amount spent on prosecution likely varies with the expected value of the patent. In fields where patents are of modest value, then, less money is likely spent on prosecution than is spent in fields where patents tend to be financial blockbusters.

54 339 US 605 (1950).

55 Id at 607–08.

56 Id at 609.

57 As this footnote itself makes clear, eevn a calim wtih literal ipemrfetcoins can siltl be reltaively esay to inertpret correctly.

These factors interact in complicated ways. For instance, a variation that, at the time of prosecution, would have been obvious to a person "skilled in the art" might seem inappropriate for protection by equivalents because applicants should be expected to anticipate obvious variations. However, courts are rightly sympathetic,[58] as the threat posed by these loopholes would indeed significantly erode patent value in the long run; applicants would in fact find it expensive to anticipate and describe every petty substitution that might be made by a strategic competitor; and there is no notice problem because competitors can easily predict that trivial variations will fall within the scope of equivalents.[59] A helpful way to think about such a case is to recognize that the doctrine of equivalents here serves to call off a wasteful arms race, a race that would otherwise encourage copyists to spend excessively on meaningless attempts to skirt literal claim language, and applicants to respond by upping the ante with respect to their attempts to craft the perfect phrase.

Viewed in light of this policy rationale, the technology-specific implications of prosecution history estoppel cut precisely backwards. Estoppel restricts equivalents most severely in cases where claim language changed significantly during the course of patent prosecution. But the factors that likely explain the high number of language changes—the difficulty the applicant faced in crafting appropriate claim language up front, and the room that was left for reasonable disagreements between applicant and examiner—suggest that these are also instances where it would have been prohibitively expensive, if not impossible, for the applicant to do better. Put simply, these are the cases that the doctrine of equivalents was designed to address. This is not to say that the doctrine should always protect such claims. But if prosecution history estoppel renders the doctrine of equivalents powerless in these cases, why have a doctrine that allows for loophole-closing at all?

The second major policy rationale that supports the use of equivalents analysis is simply the argument that there is a benefit to be gained from sometimes allowing a court to revisit patent scope even after a patent examiner has signed off on the patent's claim language. Patent prosecution takes place early in the development of a

---

58 See, for example, *Lockheed Aircraft Corp v United States*, 213 Ct Cl 395, 420 (1977) ("[E]quivalency is established where a person reasonably skilled in the art would have known of the interchangeability of an ingredient not disclosed in the patent with one that was."). But see Adelman and Francione, 137 U Pa L Rev at 697 (cited in note 51) (criticizing these cases on the ground that the patentee should have chosen better claim language).

59 Encouraging exhaustive claiming would also be counterproductive, as the resulting claim language would either contain so much detail, or be written in such generic terms, that it would be almost impossible to read.

technology, long before relevant information is available about how the invention will mature and what its economic implications will be. The patent system usually disregards this problem, trusting the applicant to know the invention well enough to craft appropriate literal claims. As applied to the most complicated and rapidly changing technologies, however, early claim drafting can be a recipe for disaster. Thus the doctrine of equivalents holds out the possibility that, in rare but appropriate circumstances, courts may in essence redraw claim boundaries using information that was not available at the time of patent prosecution.[60] Doing so has a sizeable drawback—the practice denies competitors clear notice of what is, and is not, within the patent's scope—but as applied to technologies for which the claims issued at the end of patent prosecution would otherwise regularly prove inadequate, this sort of judicial intervention is a necessary evil.

Unfortunately, as with the loophole rationale, here again the technology-specific implications of prosecution history estoppel work in reverse: in nearly every case where judicial intervention might plausibly be attractive, equivalents analysis will be disproportionately limited by estoppel. The reason is that the same lack of information that makes intervention attractive also makes it more expensive for applicants to draft comprehensive literal claim language up front, and also increases the likelihood that the examiner will disagree about the appropriate literal language and therefore require the applicant to make changes during patent prosecution. Phrased another way, the same lack of information that would make intervention attractive during litigation also will lead to the behaviors during patent prosecution that trigger prosecution history estoppel. Instead of reducing the uncertainty inherent in equivalents analysis—the justification for prosecution history estoppel invoked by the Supreme Court in *Festo*, *Graver Tank*, and *Warner-Jenkinson*—the rule of prosecution history estoppel thus again in essence repeals the doctrine.

The third policy rationale supporting the existence of the doctrine of equivalents is based on the somewhat related idea of self-selection. Few patents end up being of real economic consequence, and thus in many cases the resources invested in patent review are pure waste.[61] This is one reason why the process of patent prosecution

---

60 The doctrine of equivalents allows the court to expand patent scope. The court in addition has the power to reduce patent scope, either by invalidating a claim in full or by narrowing its scope under the seldom-used reverse doctrine of equivalents. See Merges and Duffy, *Patent Law and Policy* at 984–1003 (cited in note 50).

61 It is not always true that, where a patent ends up having no economic value, the resources invested in patent review were pure waste. For example, the process of patent prosecution might serve to reduce uncertainty by clarifying that a given patent has only narrow scope. The process might similarly sharpen claim language in a way that helps competitors successfully design around the patent. My point here is only that the resources devoted to patent evaluation

is so minimalist. It might seem odd that patent prosecution involves only the applicant and an assigned examiner, and that the average prosecution consumes a mere eighteen hours of the examiner's time.[62] But the justification is that it makes no sense to convene the National Academy of Sciences every time an inventor sees fit to file for patent protection, given that most patents spend their term gathering dust in a drawer.[63] Patents that are drawn into litigation, however, are a special subset. They have economic consequence—why else would the parties find it worthwhile to invest in litigation?—and it is therefore more likely worthwhile to invest in them the resources needed for vigorous review. This is why litigation allows parties opposed to the patent to themselves participate in the process; and this is why, instead of working with the relatively thin factual record typically cobbled together for patent prosecution, courts encourage litigants to document with care evidence regarding exactly when the patentee took each inventive step and exactly what was at each moment already known to the prior art.

Applications that are particularly attractive for this more intensive "second look" are applications for which there is reason to doubt the quality of the work done during patent prosecution. My argument, at this point predictable, is that many such applications will also be applications for which prosecution history estoppel curtails judicial discretion. The reason, as before, is that the factors that likely lead applicants to file claim language that is then altered during patent prosecution are some of the very factors that also suggest a need for the more vigorous review available through litigation. These are applications covering particularly complicated inventions or relating to rapidly developing technologies. They are therefore applications for which it is understandably difficult for applicants to predict what their assigned examiner will approve, and they are at the same time applications for which the extra firepower available in litigation would lead to more appropriate patent scope. Thus, here again, estoppel threatens to limit equivalents analysis in the core cases that equivalents analysis was designed to address. To accept a legal rule that makes equivalents disproportionately unavailable in these cases is therefore in a very

---

can sometimes be used more efficiently if they are held in reserve until more is known about which patents have economic significance.

62 Mark A. Lemley, *Rational Ignorance at the Patent Office*, 95 Nw U L Rev 1495, 1496 n 3 (2001) (noting that eighteen hours is the average total amount of "time spent reading the application, reading the submitted prior art, searching for and reading" additional prior art, and otherwise interacting with the applicant).

63 See generally id. Lemley's argument might understate the importance of removing uncertainty by declaring dud patents invalid rather than allowing them to issue, but his basic point is surely right: the limited resources spent on patent review must be allocated wisely between patent prosecution on the one hand and patent litigation on the other.

real sense to abandon the doctrine of equivalents—not directly, mind you, but through the awkward, arbitrary, and unanticipated workings of the modern prosecution history estoppel rule.[64]

C. Reforms and Responses

Concerns about differences among examiners and across technology categories can be addressed. For example, the evidentiary presumptions that under current law significantly amplify the risk of estoppel could be reversed, so that estoppel is triggered only when there is clear evidence that the applicant explicitly waived his right to argue later that a particular product or process is an equivalent. Similarly, while it might be expensive, patent examiners could be required to document more carefully the reasons for any claim language changes, again the purpose being to ensure that estoppel is triggered only in those rare instances when an applicant was indeed aware of and intending to disclaim specific coverage. These and comparable reforms would reduce the legal risk associated with amendments to claim language, and they would thereby reduce the importance of examiner and technology disparities.[65]

Admittedly, these changes would render prosecution history estoppel less effective at reducing the uncertainty inherent in equivalents analysis. The fewer times estoppel is triggered, the fewer safe harbors it creates, and thus the wider the scope of equivalents left intact. But that just confirms that estoppel is a mechanism poorly suited to the task of reducing uncertainty. Implemented conservatively, it will be triggered only rarely. Implemented moderately, it likely increases overall uncertainty by forcing patentees and their rivals to predict not only how a court will apply the doctrine of equivalents to the claims at hand, but also whether estoppel will be deemed implicated and how broad the resulting limitations will be. Implemented as it is today, with strong evidentiary presumptions papering over holes in the record, the rule is dangerously sensitive to differences among examiners and across technologies. Patent law could better improve certainty by arbitrarily suspending the doctrine of equivalents for any patent assigned a patent number that is evenly divisible by seven. From the perspec-

---

64 Even if I am wrong in all of the arguments put forward in this Part—that is, if I am incorrect in my explanation for why patterns of language alteration differ from one technology to the next, or if a reader disagrees with my interpretation of the doctrine of equivalents—note that my basic point nevertheless survives: estoppel is a rule with technology-specific implications, and those implications have been ignored in the design and implementation of the modern evidentiary presumptions.

65 Examiner differences could also be tackled head-on, perhaps by involving an additional examiner in each prosecution or by making appeals within the Patent Office more routine. The large number of patent applications filed each year, however, would likely render these reforms prohibitively expensive.

tive of patent applicants, such a rule would be equally random; and at least the divisible-by-seven rule would not disproportionately target those technologies where equivalents analysis is needed most.

CONCLUSION

I have focused in this Essay on the specific theory that prosecution history estoppel can be an effective mechanism by which to cabin the uncertainty created by the doctrine of equivalents. That theory was not targeted at random. It is the theory that the Supreme Court invoked in every case in which the Court then articulated and defended the evidentiary presumptions at issue here; and it is the theory that pervades the several opinions issued by the Federal Circuit in the context of the recent and ongoing *Festo* litigation. That said, increased predictability is not the only plausible justification for the rule of prosecution history estoppel, and the arguments I have presented here have implications for those alternatives as well.

For example, it might be that the act of negotiating claim language with an examiner puts an applicant in a better position to write clear, appropriately tailored literal claims. If so, then some form of estoppel might be an appropriate response, in essence increasing the importance of literal claim language where that language can bear the extra burden. Likewise, it might be that examiners who aggressively influence claim language are also the most conscientious about their work. If so, again estoppel would have policy allure, this time because it would obligate courts to defer more heavily to conscientious examiners. It might even be that the real motivation behind prosecution history estoppel is to encourage applicants to submit appropriately narrow claims right from the start.[66] The logic this time is that unduly broad claims are particularly likely to be changed during patent prosecution, and thus estoppel threatens most severely those applicants who claim too much in their original patent applications.[67]

These theories have strengths and weaknesses. My contribution is simply to emphasize that, no matter what the underlying policy motivation, an estoppel doctrine implemented with stringent evidentiary

---

66 See R. Polk Wagner, *Reconsidering Estoppel: Patent Administration and the Failure of* Festo, 151 U Pa L Rev 159 (2002) (developing this theory). See also John F. Duffy, *The Festo Decision and the Return of the Supreme Court to the Bar of Patents*, 2002 S Ct Rev 273, 319–20 (emphasizing that it has long been standard practice for applicants to submit unduly broad claims at first and then gradually whittle those claims down during patent prosecution).

67 There are many other plausible explanations for prosecution history estoppel. For example, estoppel might defend the integrity of Patent Office review by ensuring that an applicant cannot take one position while trying to convince an examiner to allow a claim, and then adopt a conflicting position during later litigation. Estoppel also pressures applicants to exhaust available Patent Office remedies, although query whether that is a benefit or a cost. For discussion of these and other theories, see Thomas, 47 UCLA L Rev at 204–09 (cited in note 14).

presumptions threatens two unintended consequences: it disproportionately limits the doctrine of equivalents in particular technology classes, and it makes patent value arbitrary by linking patent scope to a personal characteristic that varies considerably from one examiner to another. With respect to the uncertainty rationale, these costs are in my view devastating. As applied to other rationales, these costs are factors that must be weighed both in comparing estoppel to competing mechanisms, and in deciding the appropriate weight and direction of any evidentiary presumptions.

In short, and again no matter what the underlying theory, the rule of prosecution history estoppel must be crafted in a way that is sensitive to the practical realities of patent prosecution. The Supreme Court said as much in *Festo*,[68] yet then the Court itself failed to accomplish the task. The Federal Circuit will have the opportunity to mitigate this problem as it develops the details of the evidentiary presumptions established in *Warner-Jenkinson* and *Festo*, and as it decides in the first instance the proper presumptions to be used at other steps in estoppel analysis.[69] My purpose in this Essay is to provide the Federal Circuit with the arguments and empirical evidence it needs to engage in that process.

68 *Festo VIII*, 535 US at 738.

69 See, for example, the discussion in note 13.