

Empirical Scholarship on the Prosecution Process at the PTO

Michael Frakes and Melissa F. Wasserman

I. Introduction

Recent years have seen a surge in interest in intellectual property rights, in particular patents. This heightened level of interest has culminated in a consensus that the patent system is functioning poorly. There is widespread agreement that the United States Patent and Trademark Office (PTO or Agency) allows too many invalid patents to issue that unnecessarily drain consumer welfare and that the Agency's backlog of patent applications prevents it from providing timely review of applications. Additionally, there is growing concern that patentability decisions at the PTO are highly inconsistent across examiners—that is, the PTO's decision to grant a patent application is driven not only by the merits of the invention but also by happenstance of which examiner the application is randomly assigned.

While there is a substantial literature in economics bearing on the patent system, the administrative process by which patent rights are initially established has received scant attention. In the past decade there is a growing but nascent literature that has begun to shed empirical light on the patent examination process. This chapter will provide a brief overview of this literature, focusing only on studies that carry significant empirical components, in addition to focusing only on studies of the U.S. patent system (for purposes of tractability). Finally, we should note that the need for sound empirical guidance on the administrative process of obtaining a patent is substantial. Concerns over the patent examination process have in part prompted Congress to enact the first major patent reform act in over sixty years, spurred the Supreme Court to take a more active interest in substantive patent law, and drove the PTO to hold its first Patent Quality Summit. Yet without sufficient empirical evidence as to what features of the PTO distort the Agency's decisionmaking, policymakers have been left trying to reform the patent system without understanding the root cause of the system's pathologies.

II. Evidence on the Effectiveness of the PTO's Examination Practices: Overview

In an ideal world, patent examiners would apply the patentability standards correctly on average—i.e., generally rejecting those applications that fail to meet the patentability standards—and in a consistent manner. A number of studies have attempted to assess the PTO's success in these regards and have demonstrated shortcomings across both dimensions. Such shortcomings implicate both efficiency and equity concerns. Invalid patents unnecessarily drain consumer welfare, stunt productive research and unreasonably extract rents from innovators. In addition to invoking concerns over an inequitable process, inconsistent patentability determinations erode confidence in the PTO by creating the appearance of unfairness and arbitrariness, which could in turn diminish incentives for innovation. In the next three sections we summarize empirical studies of the patent examination process, dividing these studies into three broad categories: (1) those examining the application process more generally; (2) those demonstrating heterogeneity in application outcomes across patent examiners; and (3) those demonstrating bias toward granting by the PTO.

A. Understanding the Examination Process

There are a small but growing number of studies that have sought to assess the basic nuts and bolts of the patent application review process itself. Though these studies may fall short of identifying causal pathways between certain features of the PTO and distortions in the review process, their findings may nonetheless serve as highly valuable inputs into the broader policy debates surrounding the PTO. Before turning to these studies, this section briefly provides a background of the patent examination process.

Each year between 300,000 and 500,000 patent applications are filed at the PTO. Every patent application contains a specification, which describes the invention, and a set of claims that defines the metes and bounds of the legal rights the applicant is seeking. In addition, to satisfy applicants' duty of candor under U.S. law,¹ patent applications typically disclose to the Agency "prior art," that is previous patents, patent applications, or other publications, that are material to the patentability of the relevant invention. Before an application enters examination, it is routed to an Art Unit, a group of eight to fifteen patent examiners who review applications in the same technological field. Upon arrival, the Supervisory Patent Examiner (SPE) of that Art Unit randomly assigns the application to a specific examiner.² The assigned examiner then conducts her own prior art search to supplement that disclosed by the patent application. After completing her search, she assesses the patentability of the claims of the invention based on the criteria outlined in the Patent Act and composes a "first office action" letter to the applicant that accepts or rejects the claims. Although some applications are allowed in their entirety upon first examination, more frequently, some or all of the claims fail to meet at least one of the patentability requirements, as the examiner will detail in the first office action letter. The applicant then responds by amending the claims or disputing the rejection. After the response, a patent examiner may issue a final rejection or allow the patent to issue. Upon receiving a final rejection, an aggrieved patent applicant can restart the examination process by filing a continuation application, appealing the denied application to the Patent Trial and Appeal Board, or abandoning the application altogether.

While it may take several years from filing a patent application before hearing a final patentability decision from the Patent Office, an examiner, on average, spends only 19 hours reviewing a patent application, including reading the patent application, searching for prior art, comparing the prior art with the patent application, writing a rejection, responding to the patent applicant's arguments, and often conducting an interview with the applicant's attorney. As a result, there is considerable discussion that patent examiners are not given sufficient time to provide high quality examination. Such deficiencies are potentially problematic in light of the presumption of validity afforded to incoming patent applications—that is, if a patent examiner fails to uncover and explicitly set forth reasons as to why the application fails to meet the patentability requirements within the allotted time for examination, she is legally expected to grant the patent.

Critical to the process just summarized above is the review of prior art by the examiner in applying the patentability requirements, mainly the novelty and nonobviousness requirements. There have been several empirical studies that have examined the role of prior art in this process. These studies have

¹ This duty does not include a duty to search for material information but only a duty to disclosure material information of which an applicant is aware of.

² We conducted a series of telephone interviews with former SPEs to confirm these details of patent examination assignment. Our interviews further substantiated that SPEs do not make any substantive evaluation of an application before assigning it to a particular examiner.

exploited a 2001 change in procedure at the PTO that made it possible to measure examiner and applicant prior-art citations separately for the first time. Alcacer, Gittelman, and Sampat (2009) explored the degree to which examiners contribute to the prior art cited in issued patents. Examining all U.S. patents granted in 2001–2003 they found that examiners played a significant role in identifying prior art, adding 63 percent of all citations on average and providing all of the citations for 40 percent of patents granted. Sampat (2010) utilized the same database to examine whether applicants contribute more prior art for important inventions. Utilizing three patent-specific measures of importance for a subset of the sample patents—i.e., forward citations, renewal data, and patent family size—he finds that applicants are more likely to contribute more prior art for important inventions. Cotropia et al. (2013) built upon this early work by exploring how examiners evaluate prior art. Utilizing a 1 percent random sample of all utility patents issued in 2007 (i.e., 1,564 patents) they find that examiners overwhelmingly relied upon prior art they found themselves (rather than applicant provided) in considering whether an invention is new and nonobviousness. To the extent that examiners are not reading and applying all the art submitted to them, the Cotropia et al. analysis has several implications for patent policy. First, it casts doubt on the need for the duty of candor in the United States, which, to recap, requires patent applicants to submit prior art to the PTO. Second, these findings cast doubt on the premise underlying case law that states that prior art that was not considered by the PTO (or where there was at least no record that it was considered) likely carries more weight than prior art that had previously been considered and rejected by the PTO. That is, if examiners are not systematically reading and evaluating applicant-submitted prior art, there may be little reason to distinguish between prior art that was and was not put forth before the examiner.

Another fundamental question that arises when assessing the functioning of the review process is the degree to which patent examiners are capable of sorting meritorious from non-meritorious applications. In a recent working paper, Zhen Lei and Brian Wright confront this question by looking at a set of inventions that were patented in the United States and that likewise sought patent protection at the European Patent Office (EPO). Given that the EPO is an office with essentially similar patentability standards that invests considerably more resources in application review than the U.S., Lei and Wright follow an approach applied by other scholars and use a rejection of an application at the EPO for an invention that received a patent at the U.S. as an indication for a weak patent—i.e., as an indication for what the U.S. PTO would have done if it were to invest more heavily in its examination system. With this benchmark in mind, Lei and Wright find that claimed inventions in which U.S. examiners were able to identify a higher share of potentially citable prior art—suggesting stronger search scrutiny—were more likely to be rejected by the EPO. They interpret these findings as suggesting that U.S. examiners will exert greater effort in an attempt to invalidate an application that they feel is weak. Despite the intuitive argument that U.S. examiners may nonetheless issue invalid applications due to certain institutional features of the Agency, their analysis uses variations across a set of U.S. applications that nonetheless all issue to make an argument that U.S. examiners have the capacity to sort meritorious applications (which should be granted) from non-meritorious applications (which should be rejected).

In a forthcoming paper—which we will discuss in greater detail below—Michael Frakes and Melissa Wasserman find evidence that likewise supports the view of a rational examination process that at least has the potential to target strong applications. Frakes and Wasserman’s approach, however, generates this implication while arguably comparing the underlying validity of a set of applications that were issued in the U.S. with a set that were not, unlike Lei and Wright. As will explained further below, Frakes and Wasserman find evidence suggesting that a tightening of time constraints facing examiners may

contribute to the issuance of additional patents on the margin—i.e., the issuance of a set of patents that would have been rejected if examiners were given more time to review. They then find evidence suggesting that the patents being issued on the margin as a result of these time constraints may be of weaker underlying validity (likewise assessing validity using EPO outcomes as a benchmark) than the average issued patent—i.e., than the patents that would have likely issued even with unconstrained examination time. This suggests that U.S. examiners may target their allowances first at meritorious applications. Should pressures mount that cause them to grant additional patents on the margin, they will have to dip into the pool of marginally weaker applications. While this analysis suggests that certain features of the Agency may be leading to undesirable outcomes—as discussed further below—it does nonetheless imply a baseline degree of competency in the ability of examiners to exercise their sorting responsibilities.

In a related vein, a series of papers have sought to assess the degree to which patent examiners scrutinize applications and sort meritorious ones from non-meritorious ones by simply calculating the Agency's grant rate—that is, the probability that a given application will be allowed. This calculation is surprisingly complicated to pull off in light of several features of the examination process, including the significant length of time it takes to review applications (over which time rates of filing may change considerably) and the fact that a truly original application may ultimately lead to the filing of one or more continuation applications (creating questions as to what values should be included in the grant rate numerator and denominator). One oddity of the U.S. patent system is that an aggrieved patent applicant can always choose to start the examination process over by filing a continuation application or a request for continued examination. Early investigations into this question estimated that as much as 97% of applications are allowed by the PTO, implicating concerns of a Patent Office that is no more than a rubber stamp agency, potentially burdening society with the harms of over-granting patents delineated above (Quillen and Webster, 2001). Subsequent studies employed a number of different assumptions about the treatment of continuations and other matters in forming the relevant numerator and denominator, while also employing a range of methodological approaches—e.g., some relying on aggregate counts of various measures and some relying upon application-level data that allow one to follow individual applications throughout their life cycle. All told, the grant rates estimated in these studies vary quite a bit, falling below 50% in some studies (Quillen and Webster, 2002; Lemley and Sampat, 2008; Carley, Hedge, and Marco, 2015).

At the least, these grant-rate-calculation studies paint a picture in which the PTO is perhaps not a rubber stamp at all, but an Agency that does indeed reject a non-trivial percentage of applications. Given the findings mentioned previously of Lei and Wright and of Frakes and Wasserman (forthcoming), one might believe that those that are rejected by the PTO are indeed lacking in legal validity in the first place. Of course, these latter, largely descriptive grant-rate studies cannot determine whether the observed rates are too high or too low, and, if such rates do indeed deviate from the optimum, such studies have not causally identified any features of the PTO or of the examiners that may contribute to any such distortion. In Part V below, we turn to a summary of studies that attempt to fill this gap, including the aforementioned Frakes and Wasserman study.

It also bears emphasis that the grant rate itself provides an incomplete emphasis on the degree to which patent examiners scrutinize applications and target rewards on meritorious applications. The job of a patent examiner, after all, does not necessarily collapse to a binary decision problem of accepting or rejecting an application. They are also tasked with working with applicants to narrow the breadth of the

requested claims. There may be many initially weak applications that an examiner will ultimately allow. Focusing solely on the fact of allowance may provide a limited sense of the degree to which examiners scrutinized this questionable application and may miss any claim narrowing that the examiners were able to force. One can view any such claim narrowing as a sorting of meritorious claim breadth from non-meritorious claim breadth, even should the application itself issue in some form

The extent of claim narrowing occurring at the PTO is an underdeveloped area of scholarship, though it has been the subject of several studies to date. For instance, Mark Lemley and Bhaven Sampat (2010) find that the vast majority of patents that issue do so after undergoing some degree of amendment to the requested claims.³ In a 2004 article, Doug Lichtman tracks textual changes in the claims between the time of the application filing and the issuance of the relevant patent in order to quantify such amendments. Lichtman's analysis, however, aims to do more than that just empirically confirm that the Agency's sorting function works along this claim-narrowing margin as well; rather it primarily aims to demonstrate the degree to which patent examiners vary in how they perform this task. In this light, Lichtman joins another set of studies that seeks to explore examiner heterogeneity in practices. We turn now to a survey of this literature.

B. Heterogeneity in Patent Office Outcomes

The discussion so far has focused on understanding the process by which examiners make their decisions as well as whether examiners are capable of sorting applications based upon quality. Contemporaneous with these studies has also been the development of another line of literature that has documented substantial examiner heterogeneity in PTO outcomes. That is, these studies provide evidence that the decision to grant a patent application is driven not only by the merits of the invention but also by happenstance as to which examiner the application is randomly assigned. The existence of inter-examiner disparity raises concerns that the examination process is arbitrary as well as raising concerns as to whether examiners are systematically "missing the mark" in making validity determinations.

The seminal work of Cockburn et al. (2003) analyzes 196 examiners who had worked on 182 patents involved in United States Court of Appeals for the Federal Circuit litigation between 1997 and 2000. Evaluating all patents for which an examiner was associated, either as primary or secondary examiner, Cockburn et al. demonstrate that a notable portion of the overall variance in certain patent characteristics, including the citation received per patent examination and the citations that appear in the patents examined, can be explained simply by the identity of the examiner associated with the issued patent. Stated differently, Cockburn et al. find evidence that an applicant's outcome with the examination process is largely a function of the patent examiner that her application is randomly assigned.⁴

³ Frakes and Wasserman (2015) likewise investigate the claim narrowing role of the PTO. However, rather than quantifying the extent of claim narrowing, they find evidence in support of their prediction that the PTO issues patents of excessive breadth (as captured by the number of claims) in order to discourage the filing of continuation applications—which are costly to the Agency—during times in which the Agency is resource constrained.

⁴ This empirical observation was not especially surprising to Cockburn et al. (2003). Their extensive qualitative review of the PTO examination system revealed a process in which substantial discretion is given to patent examiners in how they deal with applications and how they search for prior art. Also, the fact that much learning within the PTO, especially before 2006, comes through a process that is akin to an apprenticeship (as

Lichtman (2004) also documents variation in the effect that examiners have on changes in language of patent claims during the application review process. Starting with a dataset of 300,000 patent applications, Lichtman, for various methodological reasons, narrowed his database down to approximately 2,200 applications which matriculated into patents. For these 2,200 patent applications Lichtman quantified the extent of textual changes between the application and the issued patent. His results suggest that differences among the assigned examiner account for about two-thirds of the variation in stringency of editing patent claims.

Various studies have built upon the findings of Cockburn et al. and Lichtman by beginning to explore certain features of the examiners that may contribute to the heterogeneity in patent examination process outcomes. Mann (2014) explored the way in which examiner tenure and experience relate to characteristics of issued patents, such as number claims, prior art references cited in issued patent, and days in examination. Analyzing over 500,000 issued patents associated with approximately 250 different examiners he finds that experience and tenure cut in opposite ways. That is, while the average number of claims and average days of examination increases with examiner experience they decrease steadily with examiner tenure.⁵ Tu (2012) examines over 1.5 million issued patents and argues that at least two distinct set of examiners exists: a group of senior examiners that issue large number of patents and a group of junior examiners that issue small number of patents. Although, it is difficult to know exactly what to make out of Tu's findings, as he does not, among other things, correct for differences in the workload expectations of examiners which increase with tenure at the Patent Office, he contends that changes in scrutiny of examiner's work product explain his findings.

Frakes and Wasserman (forthcoming 2016) examine whether the year an examiner was hired by the PTO affects how she approaches the examination process, motivated by theories of examiner learning in which initial training conditions shape persistent practice styles. In contrast to the investigations above which utilized a sampling frame of issued patents, Frakes and Wasserman use application-level data, enabling them to shed light on arguably the most important outcome of the examination process: whether the application is granted or not. Analyzing over 1.5 million patent applications filed at the Agency over a ten year period, Frakes and Wasserman find that the year a patent examiner was hired has an enduring effect on her grant rate throughout the tenure of her career. They also find that observed differences in the mean grant rates of the various examiner hiring-year cohorts align with temporal changes in both the Agency's culture regarding the allowance of patents as well as new hire training programs at the Patent Office. Their results suggest that a substantial portion of heterogeneity among examiners' application outcomes is not simply idiosyncratic to individual examiners but instead, is driven by the year in which they were hired by the Agency (combined with fluctuations over time in initial hiring conditions).

opposed to a purely centralized training system) makes it indeed possible that examiners will begin to follow divergent pathways in their examination styles.

⁵ One important limitation of Mann's analysis, however, is his treatment of time controls. The nature of the sampling frame that he employs is as such that high-experience examiners are more likely to show up in the examinations conducted later in the sample period, while low-experience examiners are likely to show up more evenly throughout the sample. This is concerning for any estimation of the link between examiner experience and examination duration to the extent that mean duration is increasingly considerably over time. To address such concerns, Mann fits linear time trends to his estimation. This control may be inadequate to the extent prosecution delays increases non-linearly over time. There may be little reason to think that duration increases linearly over time over a several decade period.

Lemley and Sampat (2012) and Frakes and Wasserman (2014), articles that will be discussed more below, examine how the experience of the examiner and the amount of examination time allocated to the examiner respectively affect patent office outcomes. Similar to Frakes and Wasserman (forthcoming 2016), Lemley and Sampat (2012) and Frakes and Wasserman (2014) also use application level data and hence are able to examine how examiner characteristics affect their decision to grant a patent. Though not the sole aim of their projects, these studies further shape our understanding of those features of examiners that contribute to the baseline heterogeneity in application outcomes across examiners.

Finally, Vishnubhakat and Rai (2015) compare patent office outcomes for Art Unit 1631 that reviews interdisciplinary inventions at the intersection of the biological and information sciences with those in matched set of applications assigned to the traditional software, Art Unit 2123. Utilizing a matched sample of approximately 60 applications in each art unit, they find that applications in Art Unit 1631 experienced significantly more rejections, particular notice-related rejections, than the conventional software applications. The authors suggest potential causal explanations including differences in the training and educational attainment of examiners in Art Unit 1631, although they acknowledge that their methodological design, along with limited data, does not enable them to make causal inferences.

C. Evidence the Agency is Deviating from its Mission

As noted above, there is general consensus that the PTO is routinely issuing invalid patents—that is, patents that are granted even though the invention fails to meet the patentability requirements. It is indisputable that invalid patents impose a multitude of costs on society. Low quality patents can be utilized by nonpracticing entities to opportunistically extract licensing revenue from innovators; they can also stunt follow on innovation. More fundamentally, invalid patents can result in supracompetitive pricing and diminished quantity without providing society with any innovative benefit.

The quality of issued patents has become such an important and visible issue that the judiciary, Congress, and the PTO have attempted to increase the quality of issued patents. Perhaps most notably, in 2011 Congress enacted the first major patent reform bill in over sixty years. The Agency was granted new adjudicatory authorities and the ability to set its own fees in an effort to improve patent quality. Despite the fact that major changes to the patent system are driven by concerns that the Agency allows too many invalid patents to issue, there is a lack of compelling empirical evidence that particular features of the system bias the PTO towards allowing patents. As a result, policymakers have been making changes to the patent system absent empirical evidence to help illuminate the actual problems at hand.

Most of the studies discussed in the previous two sections help to illuminate the examination process as well as determinants of examiner behavior. They do not, however, link their empirical analysis to particular features of the patent office that may contribute to over-granting or apply convincing strategies to causally link such features to observed granting patterns. Effective policy guidance requires analytical steps of this nature. This section now turns to discussing studies that document a pro-patent bias at the Agency.

The important work of Lemley and Sampat (2012) link one feature of the patent examination process to the outcome of whether the application is granted: the experience level (in years) of the patent examiner

randomly assigned to review the application. Looking at a sample of 10,000 applications filed in January 2001, they find a positive and monotonically increasing relationship between grant rates and the temporal experience of the examiner. They also find that as examiners gain more experience they cite less prior art, a possible indicator of lightened search scrutiny. Given that it is generally believed the European Patent Office (EPO) does a better job at screening patent applications, the authors' finding that the likelihood of the allowed US patent being rejected by the EPO (for applications filed in both the U.S. PTO and the EPO) increases with the experience of the associated US examiner suggests that the additional patents allowed by experienced examiners may be excessive in nature. Ultimately, however, Lemley and Sampat acknowledge that they cannot determine whether the senior examiners are granting excessively or whether the juniors are granting insufficiently.

While a big step forward, the major difficulty with Lemley and Sampat's analysis is its inability to attribute the positive correlation that it estimates to four possible stories: (1) that temporal experience genuinely causes patent examiners to grant at higher rates, (2) a tenure effect: that patent examiners that happen to stay at the PTO the longest and thus attain higher experience levels differ in their granting proclivities from those that stay the shortest amount of time, (3) that higher granting is due to the fact that examiners at higher pay grades—something correlated with experience in years—are given less time to review applications,⁶ and (4) a hiring-year cohort effect—the year examiners are hired (and the conditions of the PTO at that time) may have an enduring effect on an examiner's grant rate. The cross sectional nature of Lemley and Sampat's data leaves them unable to fully distinguish between the first, second and fourth stories. For instance, by comparing senior examiners and junior examiners at a given point in time, they cannot isolate a true experience effect as distinct from a selective retention effect. Moreover, Lemley and Sampat cannot distinguish between the first and third story because they do not have data sufficient to separate the experience in years of the examiner from the pay-grade level of the examiner, which is determinative of the amount of time extended to examiners.

Frakes and Wasserman (forthcoming Review of Economics and Statistics) build upon Lemley and Sampat and develop a methodological approach to separate these four stories: experience effects, tenure effects, time-allocation effects, and cohort effects. Given that their goal is to help illuminate the over-granting tendencies of the PTO, Frakes and Wasserman target their inquiry on the story out of these four that theorizes an excessive granting mechanism: the time allocated to patent examiners. The operating theory here is quite simple. Patents are assumed valid when filed. Examiners must form bases of rejection in order to reject. As such, if examiners face binding time constraints, they may reject at insufficient rates. Frakes and Wasserman test this by exploiting the fact that examination time allotments shrink as examiners are promoted within the Agency. By the time an examiner reaches GS-14, they have nearly half of the time to review applications as do comparable GS-7 examiners. Using rich data on nearly 1.5 million individual applications over a ten year period, Frakes and Wasserman track individual examiners over the course of a series of promotions that carry reductions in examination-time allocations. After performing a range of specification checks, in part, designed to rule out other explanations for the findings, the authors find that examination time becomes more constrained, patents examiners grant patents at substantially higher rates, conduct less prior art searching and perform fewer time-intensive prior-art-based rejections, mainly obviousness rejections. Under an assumption that examiners, if given sufficient time, will grant applications at an unbiased rate, the findings from this analysis suggest that

⁶ We elaborate on this time-allocation story below.

current time allocations may indeed represent binding time constraints and that they may be causing examiners to grant too many patents. Their results imply that if all examiners were given as much time to review applications as are given to GS-level 7 examiners, the PTO's grant rate would fall by 20 percent.

In the Online Appendix to their Review of Economics and Statistics article, Frakes and Wasserman further demonstrate that, of the four possible explanations listed above, the time allocation effects and hiring-year cohort effects may be most responsible for the monotonically increasing relationship documented by Lemley and Sampat between experience levels and grant rates. Higher levels of promotion within the agency contribute to higher grant rates for the reasons just stated. Moreover, Frakes and Wasserman estimate that, all else equal (e.g., even after accounting for disposition year fixed effects and other factors), examiners hired in the 1990s hold on to substantially higher grant rates throughout their careers relative to examiners hired into the 2000s,⁷ a pattern that also explains at least some degree of the baseline observation of higher grant rates associated with more years of examiner experience. After controlling for examination time allocations and hiring year effects, the relationship between examiner experience in years and grant rates surprisingly takes on an inverse-U relationship, increasing at first but then turning strongly negative.

In a series of other papers, Frakes and Wasserman empirically explore an additional set of explanations behind the PTO's perceived over-granting. In the broadest terms, these studies suggest that the PTO may respond to its highly publicized financial instability and to its substantial backlog of applications awaiting reviewing by granting more permissively to certain types of applicants, drawing on various features of the Agency that provide financial relief in the event that patents are granted. Though these papers do not purport to explain all determinants of the PTO's grant rate and do not purport to fully explain observed levels or trends in grant rates, they do suggest, at the least, that the PTO's observed grant rates may be higher than they otherwise would have been absent financial pressures.

Frakes and Wasserman's first contribution in this series targets the fee structure of the Agency. By way of quick background, the PTO has been fully user fee funded since 1991, with roughly 85% of its fee collections coming from three types of fees: (1) examination / application fees, (2) issuance fees which are collected at the time of issuance of the patent and (3) maintenance fees which are collected at 4, 8, and 12 years following the issuance of the patent in order for grantees to renew their patent protection. Even though the cost to the Agency of issuing and renewing patents is minimal, over half of the Agency's fees come from the latter "post-allowance" fees. Moreover, while the majority of the PTO's operational costs stem from examining patent applications, the examination fees are set to cover only one third of the costs incurred by the Agency to review an application. The fee structure of the Agency motivates a question: given that the majority of the PTO's operating revenue comes only in the event that they grant patents, might this cause the Agency to grant patents at elevated rates—i.e., to allow applications that they would have otherwise rejected based on an application of proper patentability standards? Frakes and Wasserman explore this question using a range of natural experiment approaches. Mainly, they test for whether the Agency began to grant at higher rates: (1) when the PTO adopted its user fee funded

⁷ In a follow-on analysis published in the Duke Law Journal, Frakes and Wasserman argue that this pattern of cohort effects is consistent with fluctuations over time in the PTO's granting culture. Agency heads promoted an atmosphere of permissive granting practices throughout the 1990s—at least to a relatively greater degree than Agency heads encouraged in the mid-2000s. The culture of the Agency at the time of hiring may imprint upon examiners as they develop enduring practices styles that persist even in the face of future swings in the PTO's philosophy regarding the application of patentability standards.

system in 1991 (relative to the pre-1991 period when it was largely funded by Congress) and thus when this incentives was potentially created in the first place, (2) when the PTO experienced aggregate financial shocks that left it with less fee revenues available to satisfy the cost of all applications awaiting review— that is when the Agency found itself in a position where it needed additional revenue and (3) when Congress, in 2004, weakened the mechanism by which it had previously diverted PTO fee collections away from PTO use, thereby making it easier for the PTO to retain its fees and thus, thereby bolstering any existing incentives it may have to elevate its fee revenues. While the authors predicted an elevation in grant rates in each instance, they further predicted that any increase in granting would be targeted on those types of applications that garner the greatest fee revenue for the Agency. Such types include large entity applicants (which pay double the fees) and applicants in technologies that consistently maintain their patents at higher rates (e.g., genetics and many information and computer technologies).

Frakes and Wasserman (2013) find that the PTO does appear to grant at higher rates to these high fee-generating types when the PTO experiences financial shocks that leave it less able to satisfy its existing examination demand with its incoming fee collections. At these same times, the PTO's backlog of patent examinations is expanding. As such, the findings suggest that the PTO may be attempting to open up resources to deal with its growing backlog by biasing its granting patterns upwards.⁸ It appears that the PTO is balancing the harms associated with elevated patenting with the harms associated with delays in its examinations that naturally occur with large backlogs. Interestingly, Frakes and Wasserman (2013) do not find that the PTO attempts to increase revenues in all circumstances, even when its financial state is healthy. They characterize the Agency as being resource-constrained but benevolent as opposed to budget maximizing. Finally, Frakes and Wasserman (2013) document a notable spike in granting proclivities (among the high fee-generating patent types) following 2004, when the Agency attained a greater ability to retain its grant rates.⁹

Frakes and Wasserman (2015) expand upon and refine this analysis in a follow-up paper published in the *Stanford Law Review*. In this subsequent analysis, they likewise theorize that in response to financial unsustainability and, concomitantly, in response to the PTO's growing backlog, they may elevate their granting rates. But now, they do not focus on features of the PTO that allow it to raise more revenues— e.g., its fee structures. Instead, they focus on an additional feature of the PTO that bears on the financial

⁸ Again, Frakes and Wasserman do not contend that these financial motivations are the only motivations driving the PTO's granting decisions; rather, they simply demonstrate that financial considerations may matter on the margin.

⁹ Frakes and Wasserman (2014) theorize that the fee structure of the PTO may, in part, be responsible for the growth of the backlog in the first place. So not only may the fee structure create an incentive to grant more to raise more money, which the PTO may act upon when it faces financial strain, but the fee structure may also contribute to that financial strain in the first place. The idea here is also straightforward. By relying so extensively on post-allowance / back-end fees to subsidize the cost of examination and by setting examination fees themselves far below the cost of examination, the PTO is exposing itself to a risk that the rate of incoming applications will grow out of step with the rate of growth of the existing stock of patents from which it garners post-allowance fees. Upon this imbalance, the PTO will not generate enough fee income to sustain that application growth and a backlog of applications will naturally ensure.

Hegde (2012) similarly theorizes that certain features of the appropriations process with Congress—e.g., the delay between the time the PTO submits its proposed budget and when it is approved by Congress—expose the Agency to risks that it will not be able to cover the expense associated with an increase in examination demand over the intervening period.

health of the Agency—its inability to finally reject a patent application as a result of mechanisms—continuation filings and request for continued examinations (RCEs)—that allow aggrieved applicants to keep pursuing their applications upon rejections by examiners. In simple terms, if applicants can keep continuing their applications, the Agency may become overwhelmed and face a piling up of applications. The Agency may attempt to avoid this outcome and discourage the use of these repeat filing devices by simply granting applications earlier in the process. Similar to the previous study, Frakes and Wasserman (2015) theorize that the PTO may target this elevated granting response on certain types of applicants over others, this time with respect to applicants in those technologies that have consistently exhibited a greater tendency to pursue continuation filings and RCE's upon rejections. They likewise find evidence that upon negative resource shocks to the Agency—or upon increases in the Agency's aggregate backlog—the PTO begins to grant at elevated rates to the targeted technologies. Consistent with the hypothesized use of such elevated granting practices to stem off continuation filings, the authors likewise find that a dip in year-to-year continuation rates for the targeted technologies at the same time.¹⁰

Finally, Jaffe and Lerner (2004) provide some evidence that the examination standards of the U.S. PTO declined in the mid-1990s but unlike the other studies discussed in this section they do not empirically identify a particular feature of the PTO that is causing the Agency to allow too many invalid patents. Jaffe and Lerner compared the growth of U.S. originated patent families, patents that are directed to the same underlying inventions that are filed and granted in multiple countries, with successful applications filed by U.S. inventors in the United States that were not part of a patent family over the time period of 1987 through 1998. They reason that if examination standards were not changing in the PTO, then successful application in the United States by U.S. inventors should grow at the same rate as U.S. originated patents that were also granted by the European Patent Office (EPO) and the Japanese Patent Office (JPO). Because they found the former grew at a rate that was twice as much as the later, they conclude the examination standards of the PTO declined.

There may of course be reasons other than declining quality of examination review that may explain diverging trends between the rate of issuance of patents at the U.S. and the rate by which U.S. originated patents issue at the EPO and JPO. For instance, perhaps there were general marketplace developments (or other developments in the innovation space) that may have encouraged relatively more filings to be directed at the U.S.P.T.O only (a possibility that is difficult to disentangle when only working with counts of issued patents). Moreover, it may also be difficult to disentangle administrative developments at the PTO that bear on how it applies patentability standards from substantive developments that reflect changes in policymakers' or judges' views over the fundamental nature of those standards themselves. This distinction is important for this chapter insofar as our goal has been to explore features of the PTO that may bias its *application* of patentability standards, whatever those standards may be—that is, our goal has been to identify reasons for the issuance of invalid patents, where invalidity is identified by deviations from the patentability requirements. EPO / JPO benchmarking as a means of assessing quality

¹⁰ The authors also demonstrate that their new analysis (Frakes and Wasserman, 2015) is empirically distinct from their prior analysis (Frakes and Wasserman, 2013). That is, they establish that an increase in grant rates among the high continuation-rate technologies cannot be explained by the fact that such technologies overlap with high fee-generating technologies. The authors note, however, when estimating for both behavioral responses simultaneously, they find that the use of elevated granting to raise additional fee revenues does not appear to materialize until the point at which Congress changes its fee diversion practices in 2004 to make it easier for the PTO to actually retain any additional fees that it collects.

of the prosecution process is useful as an exercise to the extent the legal patentability requirements are comparable across these various institutions. There may be reason to question stability in this comparability over the specific sample period employed by Jaffe and Lerner. We know that over their period of study (1987-1998), U.S. patent eligibility standards were becoming more relaxed relative to those of Japan and Europe, as the PTO for the first time enacted extensive guidelines regarding the patent eligibility of software whereas EPO and JPO continued to disallow such applications. As such, patent applicants may have chosen to file more application in the U.S. PTO than their foreign counterparts simply because more inventions became patent eligible in the U.S. than elsewhere.

III. Concluding Remarks

This chapter has provided a brief overview of the empirical literature that has explored the patent examination process. More specifically, it has focused on studies that carry a significant empirical component and that focus primarily on the U.S. patent system. Although the studies discussed in this chapter have helped to shed light on the patent application review process, including various features of the patent system that may distort the Agency's decision making, the field is still in its infancy and there is ample room for additional work in the area. This chapter will conclude by quickly delineating two fruitful areas for additional research on the patent examination process.

The studies discussed in this chapter have largely utilized either a sampling period of issued patents or application level data, the latter of which enables the exploration of what influences an examiner's decision to grant a patent. The examination process, however, is far richer than a binary grant-versus-reject story might suggest. The economic significance of a patent, and hence the potentially negative consequences of issuing an overly broad patents, depends on the patent's scope. The broader the scope of the patent, the larger number of competing products and processes that the patent can potentially block, whereas the narrower the scope of the patent, the smaller the number of products and processes that the patent can potentially block. The job of a patent examiner is not simply to decide whether or not to grant an application given the claims set forth in the application, but also to force the applicant to narrow the scope of the requested claims until the application satisfies the legal patentability standards. Understanding how various features of the patent system affect not only the examiner's decision to grant a patent but also the scope of patents issued is essential to increase the quality of patent issued by the PTO. While there are a few studies discussed above that have examined the role examiners play in narrowing the scope of claims issued, much more work is needed in this area.

More work is also needed in understanding the various levels at which decision making operates within the Agency. Many have viewed the Agency as being at the whim of the discretion of the individual examiners. The examiner heterogeneity literature discussed above certainly support this notion. However, even Cockburn et al. (2003) themselves note that much of the heterogeneity they find could be a reflection of practice styles at higher administrative units within the Agency—e.g., at the Art Unit level. Frakes and Wasserman (2013, 2014, and 2015) take a novel approach and model decision making of the Agency itself rather than the decision making of individual examiners.¹¹ Their findings support the notion that the Agency does have the ability to direct examination outcomes to serve Agency-level goals.

¹¹ To be clear, Frakes and Wasserman (2013, 2014, and 2015) recognize the import that examiner-level incentives play in PTO outcomes but examiner how features of the patent system may affect agency-level incentives.

Nevertheless, more work is needed in understanding how other levels of influence, such as mid-level managers or quality review, affect patent office outcomes.

References

Alcacer, Juan, Michelle Gittelman, and Bhaven Sampat. 2009. "Applicant and Examiner Citations in U.S. Patents: An Overview and Analysis," 38 *Research Policy* 415-427.

Carley, Michael, Deepak Hedge, and Alan Marco. 2015. "What is the Probability of Receiving a U.S. Patent," 17 *Yale Journal of Law and Technology* 203-223.

Cockburn, Iain, Samuel Kortum, and Scott Stern. 2003. "Are All Patent Examiners Equal? Examiners, Patent Characteristics, and Litigation Outcomes," in Wesley Cohen and Stephen Merrill, eds., *Patents in the Knowledge Based Economy*. National Academies Press.

Cotropia, Christopher, Mark Lemley, and Bhaven Sampat. 2013. "Do Applicant Citations Matter?" 42 *Research Policy* 844-854.

Frakes, Michael, and Melissa Wasserman. 2013. "Does Agency Funding Affect Decisionmaking?: An Empirical Assessment of the PTO's Granting Patterns," 66 *Vanderbilt Law Review* 67-148.

Frakes, Michael, and Melissa Wasserman. 2014. "The Failed Promise of User-Fees: Empirical Evidence from the U.S. Patent and Trademark Office," 11 *Journal of Empirical Legal Studies* 602-636.

Frakes, Michael, and Melissa Wasserman. 2015. "Does the U.S. Patent and Trademark Office Grant Too Many Bad Patents?: Evidence from a Quasi-Experiment," 67 *Stanford Law Review* 613-676.

Frakes, Michael, and Melissa Wasserman. Forthcoming. "Is the Time Allocated to Review Patent Applications Inducing Examiners to Grant Invalid Patents?: Evidence from Micro-Level Application Data," *Review of Economics and Statistics*.

Frakes, Michael, and Melissa Wasserman. Forthcoming. "Patent Office Cohorts," 65 *Duke Law Journal*.

Hedge, Deepak. 2012. "Funding and Performance the US Patent and Trademark Office," 30 *Nature Biotechnology* 148-150.

Jaffe, Adam and Josh Lerner, 2004. *Innovation and Its Discontents: How Our Patent System is Endangering Innovation and Progress, and What to Do About it*.

Lei, Zhen, and Brian Wright. "Why Weak Patents? Rational Ignorance or Pro-'Customer' Tilt?" Working Paper

Lemley, Mark and Bhaven Sampat. 2008. "Is the Patent Office a Rubber Stamp?" 58 *Emory Law Journal* 181-203.

Lemley, Mark and Bhaven Sampat. 2010. "Examining Patent Examination," 2010 *Stanford Technology Law Review* 2-14.

Lemley, Mark and Bhaven Sampat. 2012. "Examiner Characteristics and Patent Office Outcomes." 94 *Review of Economics and Statistics* 817-827.

Lichtman, Douglas. 2004. "Rethinking Prosecution History Estoppel," 71 *University of Chicago Law Review* 151-182.

Mann, Ronald. 2014. "The Idiosyncrasy of Patent Examiners: Effects of Experience and Attrition," 92 *Texas Law Review* 2149-2178.

Sampat, Bhaven. 2010. "When Do Patent Applicants Search for Prior Art?" 53 *Journal of Law and Economics* 399-416.

Quillen, Cecil and Ogden Webster. 2001. "Continuing Patent Applications and Performance of the U.S. Patent and Trademark Office," 11 *Federal Circuit Bar Journal* 1-21.

Quillen, Cecil and Ogden Webster. 2002. "Continuing Patent Applications and Performance of the U.S. Patent and Trademark Office –Extended," 12 *Federal Circuit Bar Journal* 35-55.

Tu, Shine. 2012. "Luck/Unluck of the Draw: An Empirical Study of Examiner Allowance Rates," 2012 *Stanford Technology Law Review* 10-55.

Vishnubhakat, Saurabh and Arti Rai. 2015. "When Biopharma Meets Software: Bioinformatics at the Patent Office," 29 *Harvard Journal of Law and Technology* ____.