R&D MANAGEMENT

Crowdsourcing in patent examination: overcoming patent examiners' local search bias

Hannes W. Lampe[1,2,*]

[1]Institute of Entrepreneurship, Hamburg University of Technology, Hamburg, 21073, Germany.
[2]Capgemini Invent, Berlin, 10785, Germany. hannes.lampe@tuhh.de

This article investigates how crowdsourcing for knowledge creation in a crucial knowledge-intense task – patent application examination – informs decision-making. It is hypothesized that patent examiners' views underly a local search bias (i.e., they rely on locally preferred and conveniently available local information), which may be overcome through crowdsourcing. To analyze this potential effect of crowdsourcing, this study analyzes USPTO's Peer To Patent initiative, opening the patent examination process to public participation for the first time. The data from this initiative is further enhanced with data from the PatentsView database and the Patent Examination Research Database. The study results provide the first empirical evidence that crowdsourcing aids a patent examination process in overcoming the examiner's local search bias – their over-reliance on internal knowledge. In particular, it is found that crowdsourcing in patent examination increases examiners' reliance on atypical and less formalized knowledge. Overall, these findings enable several theoretical and practical recommendations.

1. Introduction

In February 2018, Waymo and Uber settled a lawsuit, with Waymo being awarded Uber's shares worth $245 million. The litigation was based on two accusations: First, as predominantly communicated in the news, Waymo accused its engineers of taking more than 14,000 technical confidential files over to Uber. Second, Waymo claimed that Uber infringed four of its patents concerning Uber's laser-ranging lidar devices (BBC, 2018). However, the most surprising aspect about this litigation is that, in the same year, a not-involved engineer, Eric Swilden, filed a reexamination request with the United States Patent and Trademark Office (USPTO) (Wired, 2017). Swilden argued that cited references to the prior art (any evidence that an invention is already known[1]), which are already part of the patent were not taken into consideration by the examiner. In March 2018, the reexamination resulted in the rejection of 53 out of 56 challenged claims (USPTO, 2018).

This example elucidates the potential advantages of externals providing knowledge. Organizations' levering the increasingly distributed knowledge of marginalized actors at low cost (Jeppesen and Lakhani, 2010) is called *crowdsourcing*. However, the above example further taps into a different direction, namely that individuals focus too much on their own capabilities and the solution space known to them, called *local search bias* (March, 1991; Helfat, 1994; Hippel, 1994). Past research has established that the local search bias leads to inefficient and presumably worse decisions in various contexts that involve the evaluation of innovative activity and/or output (Martin and Mitchell, 1998; Poetz and Prügl, 2010). Using the example of the Peer-to-Patent

(PTP) initiative and its impact on patent examination processes, my work provides insights into how the local search bias can be overcome by crowdsourcing, i.e., through an open call for non-patent examiners (externals) to participate (Howe, 2006; Agerfalk and Fitzgerald, 2008; Afuha and Tucci, 2012). While crowdsourcing *per se* does not influence the likelihood of a patent being granted, it does indeed provide more atypical and less formal material to inform decision-making. By combining crowdsourcing's democracy of open participation with the legitimacy and effectiveness of administrative decision-making, this additional information may in turn serve patent examiners in making a well-thought-out decision. The study results indicate that policymakers and public officials, as well as patent examiners, are advised to consider the additional benefits that crowdsourcing, or any other form of external participation, may have on their difficult and consequential decisions alongside knowledge-intense tasks to evaluate innovative outcomes, such as, for example, the patent examination process. As such, they should wherever possible strive to increase the amount of more atypical and less formalized knowledge available. For patent applicants, being individuals or firms, it may also prove to be useful to enrich their patent submission with more, but less formalized and atypical knowledge when making their case.

While previous research has identified various benefits of crowdsourcing (Lüthje et al., 2006; Boudreau and Lakhani, 2009; Poetz and Schreier, 2012), little is known about how crowdsourcing might affect the patent application examination process. This is surprising to be noted that as a joint initiative between the USPTO and the New York Law School, PTP, which integrates externals into the patent examination process (Allen et al., 2009) has been implemented. The second pilot project has been completed, and 520 patent applications constituted this initiative. There is a serious dearth of research into the benefits of crowdsourced prior art searches, given that the first PTP initiative from the U.S. is used as a push to make Peer-to-Patent an international phenomenon; for instance, PTP has diffused to Australia, where the Queensland University of Technology collaborated with IP Australia and the New York Law School (Fitzgeral et al., 2010). Other examples of these include the Japan Patent Office (JPO), the Korean Intellectual Property Office (KIPO), and the UK Intellectual Property Office. Further, there have been repeated calls to expand the use of crowdsourcing in the patent review process (Bestor and Hamp, 2010; Ghafaele and Gibert, 2011). Although there have been some annual progress reports for these pilot projects, this study extends these reports by analyzing results beyond the first examiner action and summary statistics (Allen et al., 2008, 2009, 2012).

The remainder of this article proceeds as follows. Section 2 provides an overview of the USPTO's Peer To Patent initiative – this study's empirical focus – as well as the key weaknesses of the traditional patent application examination process. Section 3 derives the hypotheses. Next, the data are explained and models are derived to test the elaborated benefits of crowdsourcing on the patent examination process. In Section 5, the results are presented. The last section discusses the findings, outlines potential limitations and some avenues for future research, and concludes.

2. Patent examination and the Peer To Patent initiative

To analyze the PTP initiative's potential benefits, one must primarily understand the patent examination process and its downturns in its traditional form, which will be explained briefly next. An assigned examiner, usually a civil servant with a scientific or engineering background, reviews a patent application to determine whether or not the claimed invention should be granted. A patent examiner's most important task is to review the disclosure in an application and to compare it to the prior art (similar to the scientific review process). This includes reading and understanding a patent application as well as searching for prior art (in databases on granted patents, patent applications, the scientific literature, etc.) to identify whether the potential invention can contribute any further when compared to the prior art. Thus, a patent application needs to be novel and non-obvious in addition to having industrial applicability/utility for it to be granted. A patent examiner must substantially review whether a patent application complies with the legal requirements for granting a patent. In practice, this process is effortful – for instance, there may be several *office actions*[2] that require a response from inventors and their patent attorneys (USPTO, 2019).

2.1. *The shortcomings of the patent examination process*

Before elaborating on crowdsourcing's potential benefits in deriving the hypotheses, the author will first explain the key shortcomings of the traditional patent examination process. This emphasizes the importance of this study's setting and further clarifies the potential advantages of crowdsourcing in this public sector task. First, the inventors are not required to conduct prior art searches or supply the patent office with the prior art of which they lack awareness (Allen et al., 2012). Second, the patent examiners have a limited database to conduct their research: conference presentations are not available in a comprehensive form, and software code is often poorly documented and cumbersome to detect (Allen et al., 2012). In contrast to prior patents, patent examiners have restricted access to non-patent knowledge sources. Third, patent examiners have limited time to invest in a patent application. Examiners have roughly 20 hr to examine a patent and decide whether or not to grant it. Such a decision can have a huge impact, since, for instance, a 20-year grant of monopoly rights is likely to shape the future of both fundamental research and industry (Allen et al., 2012). In their examination process, examiners are expected to both digest the content of the application and prior art mentioned in this application in addition to the prior art not emphasized or cited in the application. USPTO is aware of patent examiners' lack of access to adequate information and their consequent inability to make the best decisions (USPTO Media Release, 2007). This grievance is exacerbated by examiners having limited time to conduct their prior art searches (USPTO Media Release, 2007). These downsides hamper USPTO, which is struggling with a massive backlog, as shown by an unexamined patent application inventory of more than 570,000 in January 2020 (USPTO, 2020). The next sub-chapter will now explain what the Peer To Patent initiative is and how it might counteract the downturns of the patent examination process.

2.2. *USPTO's Peer To Patent (PTP) initiative*

The PTP Initiative was launched in June 2007 and consists of an online system using Web 2.0 technology to integrate external experts into the examination process – crowdsourcing – helping examiners to identify prior art and thus support them in their prior art search. PTP's crowdsourcing process may be distinguished into five steps (displayed on the right-hand side of Figure 1): Step 1 consists of reviewing and discussing patent applications that were voluntarily submitted to PTP. Step 2 is *research and find prior art*, where members of the public conduct their research to identify reasonable prior art references which have to be mentioned or question the justification of a patent application. In step 3, prior art that is relevant to a patent application's claims is uploaded. In step 4, participants can annotate and evaluate the entirety of the submitted prior art. In the final step, the top 10 evaluated prior art references are forwarded to the USPTO to enable the examiner of an application to gain access. The patent examiner makes the final decision based on legal standards. This process combines the democracy of open participation with the legitimacy and effectiveness of administrative decision-making as displayed in Figure 1. The next chapter will now elaborate on the

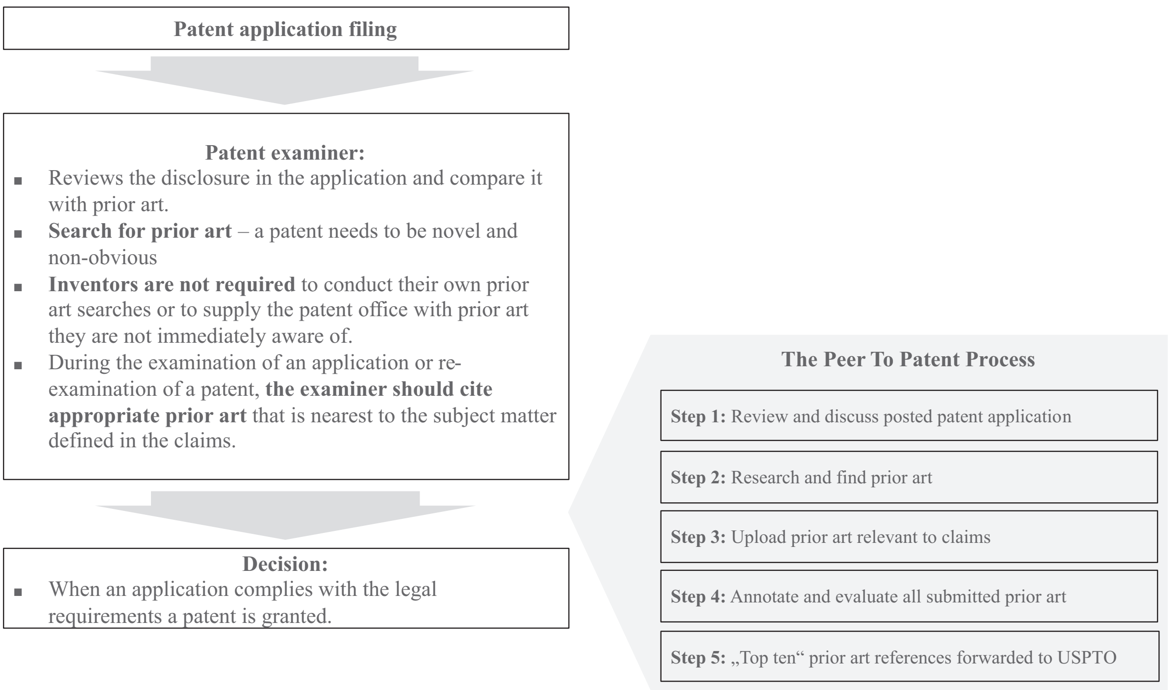

Figure 1. USPTO's peer-to-patent process.

theoretical constructs that affect this procedure and how the public participation, in the form of crowdsourcing, of PTP is likely to influence the information basis of patent examiners' decision-making.

3. Prior research and hypotheses

3.1. Prior research into patent examination

Although previous research has revealed the importance of examination quality, surprisingly, there have been very few studies on the inclusion of externals in the patent examination process. Especially, in the aforementioned example, a single engineer found evidence of prior art that was not considered by the official examiner. Three studies that have analyzed the patent examination process and the inclusion of externals stand out. First, Yamauchi and Nagaoka (2015) analyzed and found that outsourcing prior art searches lower the examination duration and likely leads to fewer appeals against examiner decisions (rejecting or granting). Similarly, Kim and Oh (2017) found that outsourcing prior art searches lowered the propensity to grant a patent and the likelihood of the reversal of an invalidation trial. However, both studies analyzed the outsourcing of prior art searches to third parties, in contrast to crowdsourcing, which enables an examiner to avoid the initial search and is depicted by a paid function subject to budgetary considerations. Crowdsourcing was analyzed in prior art searches in the form of an additional voluntary inclusion of individuals, enabling access to the wisdom of the crowd as an unpaid procedure. The third study, which closely relates to this research, was that of Kim and Mitra-Kahn (2020), which analyzed the potential effects of crowdsourcing in prior art searches. The authors focused on this procedure's unintended outcomes, showing that crowdsourcing in prior art searches led to more requests for continued examination and increased forward citations of treated patents. The authors further showed that crowdsourcing of prior art searches increased examiners' search efforts in the form of both increases in the number of search reports and the number of references added by an examiner following the first office action. Followed by elaborating on previous studies, taking the patent examination process into account, the next sub-section derives this study's hypotheses.

3.2. The effect of crowdsourcing on patent examiners' knowledge

The current research has identified that individuals residing within a single organization over-rely on internal knowledge when searching for solutions to innovation-related problems (March, 1991; Helfat, 1994; Hippel, 1994; Martin and Mitchell, 1998), further called *local search bias*. Furthermore, local search bias is a widely accepted issue in innovation-related outputs such as patents (Stuart and Podolny, 1996; Rosenkopf and Almeida, 2003). In order to overcome this bias, previous research has argued for the use of crowdsourcing. Crowdsourcing may be understood as the outsourcing of idea generation to a potentially large crowd through an open call (Howe, 2006; Agerfalk and Fitzgerald, 2008; Afuha and Tucci, 2012). Organizations that engage in crowdsourcing seek to leverage the increasingly distributed knowledge of marginalized actors at convincingly low costs (Jeppesen and Lakhani, 2010). The construct *wisdom of the crowd* (Kremer et al., 2014) is key in this case. The research has analyzed crowdsourcing in the forms of innovation contests and collaborative communities (Boudreau and Lakhani, 2009).

Crowdsourcing is an accepted tool to overcome local search bias, and thus to look beyond existing knowledge sources and tap into external sources of innovation (Lüthje et al., 2006). Some of the associated benefits include problem resolution and the capturing of value from open innovation. Here, the argumentation of tapping into external knowledge sources is in the foreground. In line with this argument, the research has shown that solutions provided by crowdsourcing outperform those developed by company experts (Poetz and Schreier, 2012). Further, Franke et al. (2014), have shown that problem-solvers from analogous markets generate more novel and less feasible solutions than solvers from the innovation task's focus market. To conclude, research has found that aggregated group solutions outperform those individuals (including experts) involved in various tasks (Budescu and Chen, 2015).

In sum, the inclusion of externals into the patent examination process – in particular, prior art searches – *via* crowdsourcing is likely to provide better outcomes, compared to internal experts' solutions (Poetz and Schreier, 2012; Franke et al., 2014) resulting in a lower likelihood of accepting patent applications. Thus, when examiners are provided with information coming from crowdsourcing, they are more likely to identify unjustified patent applications leading to higher rejection rates. Patent applications undergoing the PTP procedure are thus less likely to become granted patents.

Hypothesis 1 Crowdsourcing decreases the likelihood of patent examiners' accepting a patent application.

As argued above, local search bias might be an obstacle to the patent examination process. Furthermore, research has disentangled two causes for the occurrence of local search bias. Lakhani (2006, p. 2450) argued that "*Problem-solvers residing within a single organization will still face some level of solution myopia due to the impact of locally-preferred solution algorithms and conveniently-available local information*." Thus, according to Lakhani (2006), local search bias may occur due to two separate restrictions: (1) conveniently available information and (2) locally preferred information. The author will now elaborate on why these restrictions and, thus, a local search bias are likely among patent examiners, as well as how crowdsourcing is likely to counteract these effects.

First, the *conveniently available* information may be a reason for local search bias in the patent examination process. Prior research has argued that knowledge might be divided into formalized knowledge, expressed in specifications, objects as well as textbooks, and unformalized knowledge, depicted *via* theoretical models or models of behavior and perspectives based on empirical data and experience (Nonaka and Takeuchi, 1995). Stewart (1997) argues that formalized knowledge consists of elements such as intellectual property and databases, for instance in form of patents. Contrarily less formalized knowledge is understood as published articles or conference presentations, here depicted as non-patent references. As noted, patent examiners are restricted in accessing such less formalized knowledge (USPTO Media Release, 2007; Allen et al., 2012). For instance, examiners have access to "*some non-patent literature, they do not have the same degree of access to much of the non-patent prior art literature that exists, such as published articles, software code, and conference presentations*" (Allen et al., 2012, p. 4). Their search effort is further restricted by examiners' limited time to conduct their prior art searches (USPTO Media Release, 2007).

Research has further identified that individuals and organizations are not aware of the stock of less formalized knowledge available, and have no formalized way to access it (Du Plessis, 2007). Pyka (2002) argues that innovators seek the needed information and knowledge from professional colleagues through informal networks as valuable knowledge is often available only in less formalized formats and collaboration is a quick and efficient way to access this knowledge.

Thus, patent examiners likely have a local search bias owing to conventionally available local information in the form that less formalized knowledge is harder to access for them. As pointed out above, patents are easy to access for examiners, contrary to less formalized knowledge, in form of non-patent prior art knowledge, which is more difficult (Allen et al., 2012). Here, this research sets off the argument that crowdsourcing in patent examination enables the integration of less formalized knowledge. Thus, integrating externals into prior art searches *via* crowdsourcing enables access to additional knowledge sources that are likely not as formalized (e.g., non-patent literature). Externals may have access to additional knowledge sources (Lüthje et al., 2006) in form of less formalized knowledge for instance *via* informal networks (Pyka, 2002). By using crowdsourcing, the patent examination process, and consequently the patent examiner is likely to overcome the inherent local search bias due to conveniently available information depicted *via* formalized knowledge as patents.

Hypothesis 2 Crowdsourcing induces decision-makers to rely more on less formalized knowledge.

Second, local search bias is understood to lead individuals to overly rely on internal expertise, decreasing the probability to find alternative solutions (Lakhani, 2006). This point is crucial since examiners are likely to be specialists in a certain technological domain. For instance, Righi and Simcoe (2019) studied examiner specialization and found that examiners handle more applications from a given technology subclass or assignee than expected under a random allocation. However, specialization also means a narrower focus on certain technological subclasses – another reason for local search bias (Lakhani, 2006). Similarly, Fleming (2001) argues that individuals inevitably become narrower in their expertise as the body of knowledge expands, especially given the difficulty of searching unfamiliar domains. Again, this might be exacerbated by the time examiners have for their prior art searches (USPTO Media Release, 2007; Allen et al., 2012), which likely increases the potential effects of their local search bias. The innovation research has argued that the narrow recombination of similar knowledge can be seen as a local search (Gavetti and Levinthal, 2000; Fleming, 2001; Ethiraj and Levinthal, 2004; Kaplan and Vakili, 2015). Thus, examiners seem well aware of prior art in the same technology class as the examined patent. Further, examiners are likely to be aware of the prior art typically referenced in their specialized domain. However, examiners are less aware of atypically referenced technology classes, resulting in a local search bias that favors references typically associated with technology classes.

Via crowdsourcing, examiners are likely to overcome their locally preferred solutions by including more atypical knowledge. Having a broader audience identifying prior art is likely to include a wider range of prior knowledge, which is likely to come from technological domains typically not associated with the focal patent's technology class. Thus, *via* the PTP, examiners' references' technology classes will become more atypical:

Hypothesis 3 Crowdsourcing induces decision-makers to rely more on atypical knowledge.

4. Method and data

The above Hypotheses focus on two different aspects of the examiner's decision-making: Hypothesis 1 focuses on the overall likelihood to grant a patent, whereas Hypothesis 2 and 3 proposed two information characteristics: formalized and atypical knowledge. Thus, two separate datasets had to be built to test these hypotheses.

Dataset 1 tested for a higher likelihood to grant a patent application (H1). To examine if PTP-treated patent applications have a higher likelihood to become granted patents, this dataset took patent applications into account. The Patent Examination (PatEx) Research Database, sourcing data from the Public Patent Application Information Retrieval system (Public PAIR) (Graham et al., 2015) was used to build dataset 1. The original Public PAIR dataset contained information on 11,125,755 patent applications, which was then matched with the PTP-treated applications (this information was downloaded from the PTP homepage – https://www.peertopatent.org – and included 520 patent applications that underwent the PTP). Using the United States Patent Classification (USPC) technology classes (36) and the associated filing year (2005 to 2011) of each application in the PTP sample, all applications were further subsetted, matching these filing year/technology class combinations, which resulted in 375,532 patent applications. Again, the dataset comprises patent applications with the same technology classes and filing years as the PTP-treated patent applications, not blurring the results with effects associated due to different acceptance rates over the years or technology classes. Deleting applications without a final decision status (granted or not granted) resulted in 338,051 patent applications (of which 520 were part of the PTP initiative). This dataset was then used to test H1.

To build the second dataset, testing H2 and H3, which focus on information characteristics – formalized and atypical knowledge – of patent examiners, additional information on references made by the examiner was needed. This information is given for granted patents by the PatentsView database (Leydesdorff et al., 2017). Similar to the building of dataset 1, here granted patents with the same technology class and filing year as the PTP-treated patents were included. Owing to focusing only on granted patents, the observations dropped to 221,148.[3] In addition – as the focus here lies on the examiner references – the dataset was further subsetted to granted patents examined by the same examiners as those that underwent the PTP treatment, which resulted in 6806 observations.

4.1. *Case–control samples*

Case–control samples are used to restrict or subset the dataset based on a certain restriction. In addition to the base case (explained above for each dataset) information on patent's abstracts and patent citations was used. Unfortunately, these were only available for granted patents, and thus only the case–control of the second dataset was altered. In addition to base restrictions (technology/year classification), only those patents with the same number of forwarded citations (as the PTP-treated ones) were kept to allow for differences due to a patent's value and technological success (Fleming, 2001; Jung and Lee, 2016).

Further, the topic modeling (Blei et al., 2003; Kaplan and Vakili, 2015) was used to build another case–control sample. Topic modeling allows one to uncover automatically identified topics[4] and themes that are latent in a collection of documents and to detect which theme composition best accounts for each document. The algorithm then uses this information to detect document similarities. This information was then used to subset the dataset to those patents that underwent the PTP procedure and their most similar (based on their abstracts) counterparts (not part of the PTP program) enabling a robust control sample (Arts et al., 2018). This sample consisted of 582 granted patents.

4.2. *Dependent and independent variables*

To test the hypotheses, three models were built, each with a different dependent variable. Hypothesis 1 (based on dataset 1) focused on the likelihood of a patent application becoming a granted patent. To test whether a PTP-treated patent application is less likely to be granted, a dummy variable, indicating if a patent application was *granted* or not, was used.

Hypothesis 2 assumed that crowdsourcing induces decision-makers to rely more strongly on less formalized knowledge. To measure if the PTP initiative led to higher use of less formalized knowledge, the number of non-patent references in a granted patent was used. In particular, the total *nonpatent references* as a proxy for those made by the examiner were used, since no further information on these reference types was given. However, when the non-patent references increase, the share of nonpatent references made by an examiner may also increase.

Hypothesis 3 proposed that PTP-treated patents show more atypical references made by the examiner. To test this hypothesis, the author constructed a variable that measures how atypical references made by the examiner were. Following the previous research and arguing that examiners are specialized and experienced in certain technology classes (Righi and Simcoe, 2019), to measure the atypicality of examiner references based on the referenced patent's technology class compared to the focal patent's technology class. Atypical technology class recombinations (comparison of cited and focal patent technology class) by examiner added references of granted patents were used as the third dependent variable. This measure was built by adopting and modifying a measure from Lo and Kennedy (2015) in two steps. First, the proximity index for each pair of technology classes i and j was computed. This index measures how typical two technology classes are in their combination. c depicts the focal patent's primary technology class, while j refers to the primary technology class of a prior art patent reference added by an examiner. The proximity index P_{ij} for two classes i and j is the average of two proportions; thus, it is calculated by dividing C_{ij}, by the total number of times a technology class is mentioned in the entire sample (C_i) and the reference's technology class is referenced (C_j) to, respectively. C_{ij} is the number of referenced (by examiner) patents from technology class j in technology class i patents and thus depicts the number of times one technology class is cited in another technology class (aggregated over patents). This can be written as follows:

$$P_{ij} = \frac{1}{2}\left(\frac{C_{ij}}{C_i} + \frac{C_{ij}}{C_j}\right) \tag{1}$$

where C_{ij} is the number of times a patent of classes i had an examiner reference a patent with class j. C_i are the total number of times that class i appeared as the primary technology class in patents, and C_j depicts the total number of examiner references to a patent with primary technology class j in a certain time span. For the base case, the focal patent's filing year as the time span was used. For additional robustness testing, this measure was built using 1-year, 3-year, and 5-year time windows. For instance, a 3-year time window takes not only citations of patents in the focal year into account, but further includes citations from patents applied for in the previous three years to measure the typicality of examiner references.

Second, to measure how typical an examiner reference is, compared to the entirety of all examiner references, the proximity indices for every pair of classes were averaged. Larger scores imply that a class combination occurs together more often (and is thus more typical), while smaller scores mean a more unusual or atypical blend of technology class combinations. Typicality was measured as follows:

$$T = \frac{\sum P_{ij}}{[L(L-1)]/2} \tag{2}$$

Here, T is a patent examiner referencing typicality concerning the focal patent's technology class. P_{ij} are the measures of proximity for each pair of classes, and thus, the focal patent's technology class and the examiner-referenced patent's technology class. L is the number of the focal patent's references coming from the examiner. To measure atypicality, this variable was then subtracted from 1, resulting in the dependent variable *atypicality of examiner references*. A more extensive derivation of this atypicality measure can be found in Lo and Kennedy (2015).

4.3. Independent and control variables

The unique feature of the PTP initiative is that this treatment enables the comparison of the patents that underwent the PTP procedure with those that did not. To test this study's hypotheses, the focal independent variable is therefore a dummy variable, indicating one if a patent (application) was part of the PTP procedure and zero otherwise. All three models included dummies for the *filing year* and the *technology class* of an application or patent to control for differences in the dependent variable due to these two factors. Since recent research has identified the key role of examiner specialization and experience (Righi and Simcoe, 2019), a measure for *examiner experience* was included in all models. This measure depicts the number of years an examiner has been examining patent applications related to the filing year of the focal patent or patent application. In Models 2 and 3, which considered granted patents, the total *number of claims* of a patent was further used (Kaplan and Vakili, 2015; Jung and Lee, 2016) (since this information was only available for patents, not applications). Model 3 further included controls for the total *number of references* (Fleming, 2001) and the *number of non-patent references* (Jung and Lee, 2016).

4.4. Estimation methods

To test the hypotheses, three dependent variables were used. All of these variables differ in their characteristics and distributions, and these differences were taken into account by using different estimation approaches. Model 1, which tests H1, used a dummy variable as a dependent variable, leading to the use of logistic regression (Maddala and Lahiri, 1992). To test H2, this study incorporated the number of non-patent references, a typical count variable that cannot assume values smaller than zero and corresponds to counts (Lampe and Reerink, 2021). Following Hausman et al. (1984), the obvious approach would be to use a Poisson model. However, additional tests showed that this variable is overdispersed. This led to the use of a negative binomial model. Third, the testing of a patent's reference atypicality was expressed *via* a continuous variable following a normal distribution, which led to an ordinary least squares (OLS) regression to test the predicted effect. The literature has identified the examiners' key role in several aspects of patents (see Lemley and Sampat, 2012; Righi and Simcoe, 2019). To account for the importance of examiners and to allow for serial autocorrelation, clustered standard errors were used, treating each examiner as a cluster in all models. Table 1 presents the descriptive statistics and a correlation matrix for the variables used in Models 2 and 3.

5. Results

Table 2 presents the final regression results. Model 1 shows the results of the logistic regression in combination with the patent application dataset to test for H1. Models 2 and 3 used the final dataset on granted patents. Model 2 used a negative binomial regression model. Model 3 tested H3 and used an OLS regression.

Model 1 tested for H1, predicting that a PTP-treated patent application is less likely to be granted. The PTP treatment's effect was insignificant; thus, H1 was not supported. Previous research was also unable to show a significant relationship between crowdsourced prior art searches and the likelihood of a patent application is granted (Kim and Mitra-Kahn, 2020). H2 predicted that a PTP-treated patent is likely to have more non-patent references, and thus a higher use of less formalized knowledge. Model 2 showed that this effect was significant ($\beta = 0.342$, $P < 0.01$); thus, H2 was supported. Model 3 tested H3. The proposed positive effect of the PTP treatment on the atypicality of examiner prior art references ($\beta = 0.061$, $P < 0.05$) was supported.

To challenge the current study's findings' robustness, several robustness tests were conducted. First, to test H3, the dependent variable was determined not only with respect to the focal patent's filing year, but further rebuilt the variable using 1-year, 3-year, and 5-year time windows. The results were qualitatively consistent. Further, the matching procedure was challenged to build the case–control sample. Then, PTP-treated patents (again building on the filing year/technology class combinations) were matched with those with the same number of forwarded citations to allow for differences due to a patent's value and technological success (Fleming, 2001; Jung and Lee, 2016). The results based on this matching procedure appear in Table 3 (models 1 and 2). Further, I used topic modeling (Blei et al., 2003; Kaplan and Vakili, 2015) to build another case–control sample, consisting of 582 granted patents. This dataset was then used to test H2 and H3. The results are displayed in models 3 and 4 in Table 3. Overall, the results remained the same. Only for model 4, which tested H3, was a drop in the significance level observed. This may be due to the drop in the number of observations.

Table 1. The descriptive statistics

	Mean	Mean	SD	1	2	3	4	5
1	Nonpatent references	8.67	26.54					
2	Atypicality of examiner references	0.76	0.40	0.03				
3	PTP-treated (dummy)	0.05	0.21	0.04	0.02			
4	Examiner experience (years)	11.94	10.85	−0.01	0.07	−0.05		
5	Number of claims	18.26	10.25	0.14	0.03	−0.03	0.01	
6	Number of references	17.50	38.32	0.61	0.10	0.01	0.05	0.11

Table 2. The main results: testing the hypotheses

	Dependent variable		
	Granted patent	Nonpatent references	Atypical references
	Logistic	Negative binomial	OLS
	(1)	(2)	(3)
PTP treatment (dummy)	−0.006	0.342***	0.061**
	(0.099)	(0.098)	(0.028)
Examiner experience (in years)	0.052***	0.029***	0.003
	(0.001)	(0.006)	(0.002)
Number of claims		0.027***	0.0005
		(0.002)	(0.001)
Number of references			0.001***
			(0.0002)
Number of nonpatent references			−0.0002
			(0.0003)
Constant	0.379***	1.982***	0.700***
	(0.077)	(0.291)	(0.093)
Year dummies ($n=7$)	Yes	Yes	Yes
Technology class dummies ($n=36$)	Yes	Yes	Yes
Observations	338,051	6,806	4,231
Unique examiners	4,511	247	228
R^2			0.095
Adjusted R^2			0.086
Theta		0.372***	
		(0.007)	
F-statistic			10.420***

Robust standard errors in parentheses clustered at the examiner level.
*$P<0.1$, **$P<0.05$, ***$P<0.01$.

6. Discussion and conclusion

Previous research has shown that the local search bias (March, 1991; Helfat, 1994; Hippel, 1994; Martin and Mitchell, 1998) leads to less efficient and probably worse decisions in various fields. The introductory example, where a single engineer filed for the reexamination of a patent that two global companies were fighting over, emphasized how local search bias might lead patent examiners to rely on locally-preferred solution spaces and conveniently-available local information. However, the example further elucidates that externals might possess knowledge that may be key for decision-making. Here, this study sets off, providing insights into how the local search bias can be overcome by the integration of externals *via* crowdsourcing (Howe, 2006; Agerfalk and Fitzgerald, 2008; Afuha and Tucci, 2012) to alter the information basis of decision-makers. Using the example of the PTP initiative, it has been demonstrated that crowdsourcing provides more atypical and less formalized knowledge to inform decision-making, which may assist the patent examiner to make well-informed decisions. However, crowdsourcing does not *per se* influence the likelihood of a patent being granted. The results of this study indicate that a variety of decision-makers: patent examiners, public officials, as well as policymakers, are advised to consider the additional benefits of crowdsourcing, as well as other forms of external participation, to alter the decision basis for knowledge-intensive tasks evaluating innovative outcomes (in the current case, the patent examination process). Thus, decision-makers should whenever possible try to enhance less formal and more atypical knowledge available.

These findings extended the research into local search bias (March, 1991; Helfat, 1994; Hippel, 1994; Martin and Mitchell, 1998; Poetz and Prügl, 2010). As Fleming (2001) argued, individuals, become narrower in their expertise as the body of knowledge expands, especially given the difficulty of searching unfamiliar domains. In the current context, crowdsourcing yields access to

Table 3. Robustness tests

	Dependent variable			
	Nonpatent references	Atypical references	Nonpatent references	Atypical references
	Negative binomial	OLS	Negative binomial	OLS
	Citation matching		Examiner and text-based matching	
	(1)	(2)	(3)	(4)
PTP treatment (dummy)	0.403***	0.061**	0.306***	0.070*
	(0.094)	(0.030)	(0.111)	(0.042)
Examiner experience (in years)	0.010***	0.002***	0.049***	−0.002
	(0.001)	(0.0004)	(0.017)	(0.007)
Number of claims	0.022***	0.0003	0.024***	−0.001
	(0.001)	(0.0002)	(0.007)	(0.003)
Number of references		0.001***		0.001
		(0.0001)		(0.001)
Number of nonpatent references		−0.001***		0.0001
		(0.0001)		(0.001)
Constant	1.255***	0.807***	0.331	0.578*
	(0.142)	(0.044)	(0.781)	(0.306)
Year dummies ($n=7$)	Yes	Yes	Yes	Yes
Technology class dummies ($n=36$)	Yes	Yes	Yes	Yes
Observations	86,271	41,819	582	357
Unique examiners	3,401	2,927	238	181
Theta	0.384***		0.614***	
	(0.002)		(0.038)	
F-statistic		80.969***		1.768***

Robust standard errors in parentheses clustered at the examiner level.
*$P<0.1$, **$P<0.05$, ***$P<0.01$.

more atypical knowledge, countering this effect. Furthermore, even though less formalized knowledge plays a more dominant role in the innovation process (Du Plessis, 2007), individuals and organizations neither are aware of the stock of less formalized knowledge nor have a formalized way to access it (Du Plessis, 2007). In addition, this study's results show that *via* crowdsourcing less formalized knowledge is added to the information basis of decision-makers. Thus, crowdsourcing is a vital aspect to get access to additional less formalized, and more atypical knowledge for the decision-making process. Furthermore, this study's results increase our understanding of crowdsourcing (Howe, 2006; Agerfalk and Fitzgerald, 2008; Afuha and Tucci, 2012), by taking the two knowledge characteristics of less formalized and more atypical knowledge into account. The finding that crowdsourcing enables access to less formalized and more atypical knowledge is further likely consistent in other settings when searching for solutions to innovation-related problems as these are depicted *via* similar characteristics (March, 1991; Helfat, 1994; Hippel, 1994; Martin and Mitchell, 1998).

Several managerial implications might be derived from this study's results. Crowdsourcing provides more atypical and less formalized knowledge to inform decision-making. Where, in public sector entities (such as the USPTO), enabling access to additional databases may be a highly bureaucratic task, externals may have this access and may well be willing to share their knowledge. This holds true, especially for less formalized knowledge, identified as being hard to access (Pyka, 2002; Du Plessis, 2007). For instance, conference presentations are not easily accessible in full. Additional software source code is hard to identify as such and is even harder to access. Thus, the PTP initiative incorporated not only externals' heterogeneous knowledge but also the additional knowledge sources they have access to into the patent application examination process. In sum, several sources of additional knowledge may well be accessed, and there are most likely time savings owing to externals having evaluated prior art references prior to submitting these to an examiner. Managers are thus well advised to use crowdsourcing to increase atypical and less formalized knowledge for their decision-making.

Further, I have added to the research into the patent examination process, particularly prior art searches, a topic that has drawn very little research attention (Yamauchi and Nagaoka, 2015; Kim and Oh, 2017; Kim and Mitra-Kahn, 2020). In contrast to most previous research, by analyzing the full outsourcing of prior art searches (Yamauchi and Nagaoka, 2015; Kim and Oh, 2017) and thus a paid function, I have extended the research by focusing on the voluntary crowdsourcing of this search procedure. It has been shown that information gathered *via* crowdsourcing adds to patent examiners' decision-making basis by providing more atypical and less formalized knowledge. To my best knowledge, only one empirical study has focused on the crowdsourcing phenomenon; however, Kim and Mitra-Kahn (2020) did not analyze *how* and *if* the PTP initiative positively affects the knowledge accessed by an examiner, and thus their decision-making basis and its underlying information. Thus, this study is the first to prove that crowdsourcing in the patent examination process provides more atypical and less formal material to inform decision-making.

It is well known that examiners have limited time to conduct prior art searches and that they can make the ideal decision with access to adequate data (USPTO Media Release, 2007). Adding more atypical and less formal knowledge could be the first building block to improve the patent examiner's underlying information basis for decision-making. Furthermore, the PTP initiative is a push, to be diffused toward other countries, increasing the necessity to understand its benefits better. However, an important area for future research might be the comparison of the PTP initiative, and thus crowdsourcing in the patent examination process, with automated approaches to inform decision-making. For instance, machine learning, to identify prior knowledge, might yield extensive advantages. However, although machine learning is an ever-advancing research field, with a high yield of benefits in several areas, data availability could be an obstacle only to be overcome by the PTP initiative. In addition, PTP is relatively easy to implement in the patent examination process. Thus, the benefits of PTP seem to outweigh the downturns.

Previous research has shown how prone patent examinations' outcomes are to increased examiner workloads (Kim and Oh, 2017) or even different weather conditions (Kovács, 2017). These downsides are apparent in USPTO's massive backlog of more than 570,000 unexamined patent applications in January 2020 (USPTO, 2020). Further, the patent examination process, and thus its outcomes have far-reaching consequences, since a patent that grants a 20-year monopoly right is likely to shape the future of fundamental research as well as an industry. In line with this argumentation, it has been shown how the public, *via* the crowdsourcing of prior art searches, may contribute to this important public sector task by providing more atypical and less formalized knowledge for decision-making. These findings have further importance since there have been repeated calls for the crowdsourcing of prior art searches (Bestor and Hamp, 2010; Ghafaele and Gibert, 2011) yet very little empirical evidence. Further, the first PTP initiative from the U.S. is used as a push to make Peer-to-Patent an international phenomenon, including the Japan Patent Office (JPO), the Korean Intellectual Property Office (KIPO), and the UK Intellectual Property Office. Thus, this study is among the first to enable the verification of this procedure's benefits, especially in which contexts crowdsourcing may yield benefits.

While this study made a thorough effort to analyze the potential effects of crowdsourcing prior art searches on examiner information for decision-making, this study has its own limitations. The study findings require external validity discussion. In some cases, the limitations offer fruitful future avenues for research. The specific national context of the U.S. has been the primary focus.; this has allowed me to empirically analyze the first PTP-like initiative but has omitted similar approaches' effects in different national contexts, limiting the findings' generalizability. Thus, it would be interesting to test the proposed effects in other national contexts if such initiatives have been conducted. Owing to data availability, the H2 and H3 were tested based on granted patents. Deeper insights, especially concerning individual information on references and identifying whether references arise from the PTP initiative, would improve the results with a fruitful avenue for future research. Since non-patent references are not identified by whom these are cited, the total number was used as a proxy for those made by the examiner. However, more detailed information could improve our understanding of the benefits of the PTP initiative.

Data availability statement

The empirical analysis uses data from USPTOs Peer To Patent (PTP) initiative (n = 520) and builds two new datasets (including control samples) by adding two additional data sources: Public PAIR (resulting in

338,051 observations) and PatentsView (resulting in 6,806 observations). All of these datasets are openly available from: https://www.peertopatent.org, https://www.patentsview.org/download/, https://www.uspto.gov/learning and resources/electronic data products/patent examination research dataset public pair.

Acknowledgement

Open Access funding enabled and organized by Projekt DEAL.

References

Afuha, A. and Tucci, C. (2012) Crowdsourcing as a solution to distant search. *Academy of Management Review*, **37**, 3, 355–375. https://doi.org/10.5465/amr.2010.0146.

Agerfalk, P. and Fitzgerald, B. (2008) Outsourcing to an unknown workforce: exploring opensourcing as a global sourcing strategy. *MIS Quarterly*, **32**, 2, 385–409. https://doi.org/10.2307/25148845.

Allen, N., Casillas, A., Chichetti, S., DeFrances, M., Kabir, T., Segro, C., and Webbink, M. (2012) *Peer to Patent: First Pilot Final Results*. New York, NY: Center for Patent Innovation, New York Law School. https://www.peertopatent.org/wp-content/uploads/sites/2/2013/11/First-Pilot-Final-Results.pdf.

Allen, N., Casillas, A., Deveau-Rosen, J., Kreps, J., Lemmo, T., Merante, J., Murphy, M., Osowski, K., Wong, C., and Webbink, M. (2009) *Peer to Patent: Second Anniversary Report*. New York, NY: Center for Patent Innovations, New York Law School.

Allen, N., Ingham, J., Johnson, B., Merante, J., Noveck, B., Stock, W., Tham, Y., Webbink, M., and Wong, C. (2008) *Peer to Patent: First Anniversary Report*. New York, NY: Center for Patent Innovations, New York Law School.

Arts, S., Cassiman, B., and Carlos, J. (2018) Text matching to measure patent similarity. *Strategic Management Journal*, **39**, 62–84. https://doi.org/10.1002/smj.2699.

BBC. (2018) *Uber Settles with Waymo on Self-Driving*. https://www.bbc.com/news/technology-43010348

Bestor, D. and Hamp, E. (2010) Peer to patent: a cure for our ailing patent examination system. *Northwestern Journal of Technology and Intellectual Property*, **9**, 16–28. https://scholarlycommons.law.northwestern.edu/njtip/vol9/iss2/1.

Blei, D.M., Ng, A.Y., and Jordan, M.I. (2003) Latent dirichlet allocation. *Journal of Machine Learning Research*, **3**, 993–1022. https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf?TB_iframe=true&width=370.8&height=658.8.

Boudreau, K.J. and Lakhani, K.R. (2009) How to manage outside innovation. *MIT Sloan Management Review*, **50**, 4, 69–76. https://moodle2.units.it/pluginfile.php/186278/mod_resource/content/1/RD_4_How_to_manage_Outside_Innovation.pdf.

Budescu, D.V. and Chen, E. (2015) Identifying expertise to extract the wisdom of crowds. *Management Science*, **61**, 2, 267–280. https://doi.org/10.1287/mnsc.2014.1909.

Du Plessis, M. (2007) The role of knowledge management in innovation. *Journal of Knowledge Management*, **11**, 4, 20–29.

Ethiraj, S. and Levinthal, D. (2004) Modularity and innovation in complex systems. *Management Science*, **50**, 2, 159–173. https://doi.org/10.1287/mnsc.1030.0145.

Fitzgeral, B.F., McEniery, B.J., and Ti, J. (2010) *Peer to Patent Australia: First Anniversary Report*. Faculty of Law at Queensland University of Technology. https://eprints.qut.edu.au/39350/.

Fleming, L. (2001) Recombinant uncertainty in technological search. *Management Science*, **47**, 1, 117–132. https://doi.org/10.1287/mnsc.47.1.117.10671.

Franke, N., Poetz, M., and Schreier, M. (2014) Integrating problem solvers from analogous markets in new product ideation. *Management Science*, **60**, 4, 1063–1081. https://doi.org/10.1287/mnsc.2013.1805.

Gavetti, G. and Levinthal, D. (2000) Looking forward and looking backward: cognitive and experiential search. *Administrative Science Quarterly*, **45**, 1, 113–137. https://doi.org/10.2307/2666981.

Ghafaele, R. and Gibert, B. (2011) Crowdsourcing patent application review: leveraging new opportunities to capitalize on innovation? *Intellectual Property Quarterly*, **3**, 23–33. https://mpra.ub.uni-muenchen.de/id/eprint/38092.

Graham, S., Marco, A., and Miller, R. (2015) The USPTO patent examination research dataset: a window on the process of patent examination. *USPTO Economic Working Paper No. 2015-4*. https://ssrn.com/abstract=2702637.

Hausman, J.A., Hall, B.H., and Griliches, Z. (1984) Econometric models for count data with an application to the patents-R&D relationship. *Econometrica*, **52**, 4, 909–938. https://doi.org/10.2307/1911191.

Helfat, C. (1994) Firm-specificity in corporate R&D. *Organization Science*, **5**, 173–184. https://doi.org/10.1287/orsc.5.2.173.

Hippel, V. (1994) "Sticky information" and the locus of problem solving: implications for innovation. *Management Science*, **40**, 4, 429–439. https://doi.org/10.1287/mnsc.40.4.429.

Howe, J. (2006) The rise of crowdsourcing. *Wired Magazine*, **14**, 6, 1–4. http://www.wired.com/wired/archive/14.06/crowds_pr.html.

Jeppesen, L. and Lakhani, K. (2010) Marginality and problem-solving effectiveness in broadcast search. *Organization Science*, **21**, 5, 1016–1033. https://doi.org/10.1287/orsc.1090.0491.

Jung, H.J. and Lee, J.J. (2016) The quest for originality: a new typology of knowledge search and breakthrough inventions. *Academy of Management Journal*, **59**, 5, 1725–1753. https://doi.org/10.5465/amj.2014.0756.

Kaplan, S. and Vakili, K. (2015) The double-edged sword of recombination in breakthrough innovation. *Strategic Management Journal*, **36**, 10, 1435–1457. https://doi.org/10.2139/SSRN.2261731.

Kim, J.-H. and Mitra-Kahn, B. (2020) The unintended consequences of crowdsourcing prior art search. *Applied Economics*, **52**, 24, 2569–2579. https://doi.org/10.1080/00036846.2019.1693022.

Kim, Y. and Oh, J. (2017) Examination workloads, grant decision bias and examination quality of patent office. *Research Policy*, **46**, 5, 1005–1019. https://doi.org/10.1016/j.respol.2017.03.007.

Kovács, B. (2017) Too hot to reject: the effect of weather variations on the patent examination process at the United States patent and trademark office. *Research Policy*, **46**, 10, 1824–1835. https://doi.org/10.1016/j.respol.2017.08.010.

Kremer, I., Mansour, Y., and Perry, M. (2014) Implementing the "wisdom of the crowd". *Journal of Political Economy*, **122**, 5, 988–1012. https://doi.org/10.1086/676597.

Lakhani, K.R. (2006) Broadcast search in problem solving: attracting solutions from the periphery. In: *2006 Technology Management for the Global Future-Picmet 2006 Conference*. IEEE, pp. 2450–2468. https://doi.org/10.1109/PICMET.2006.296838.

Lampe, H.W. and Reerink, J. (2021) Know your audience: how language complexity affects impact in entrepreneurship science. *Journal of Business Economics*, **91**, 1025–1061. https://doi.org/10.1007/s11573-020-01027-4.

Lemley, M.A. and Sampat, B. (2012) Examiner characteristics and patent office outcomes. *The Review of Economics and Statistics*, **94**, 3, 817–827. https://doi.org/10.1162/REST_a_00194.

Leydesdorff, L., Kogler, D., and Yan, B. (2017) Mapping patent classifications: portfolio and statistical analysis, and the comparison of strengths and weaknesses. *Scientometrics*, **112**, 1573–1591. https://doi.org/10.1007/s11192-017-2449-0.

Lo, J.Y.-C. and Kennedy, M.T. (2015) Approval in nanotechnology patents: micro and macro factors that affect reactions to category blending. *Organization Sicence*, **26**, 1, 119–139. https://doi.org/10.1287/orsc.2014.0933.

Lüthje, C., Herstatt, C., and von Hippel, E. (2006) User-innovations and "local" information: the case of mountain biking. *Research Policy*, **34**, 6, 951–965. https://doi.org/10.1016/j.respol.2005.05.005.

Maddala, G.S. and Lahiri, K. (1992) *Introduction to Econometrics*. Volume **2**. New York: Macmillan.

March, J. (1991) Exploration and exploitation in organizational learning. *Organization Science*, **2**, 1, 71–87. https://www.jstor.org/stable/2634940.

Martin, X. and Mitchell, W. (1998) The influence of local search and performance heuristics on new design introduction in a new product market. *Research Policy*, **26**, 753–771. https://doi.org/10.1016/S0048-7333(97)00037-1.

Nonaka, I. and Takeuchi, H. (1995) *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*. New York: Oxford University Press.

Poetz, M. and Prügl, R. (2010) Crossing domain-specific boundaries in search of innovation: exploring the potential of pyramiding. *Journal of Product Innovation Management*, **27**, 897–914. https://doi.org/10.1111/j.1540-5885.2010.00759.x.

Poetz, M. and Schreier, M. (2012) The value of crowdsourcing: can users really compete with professionals in generating new product ideas. *Journal of Product Innovation Management*, **29**, 245–256. https://doi.org/10.1111/j.1540-5885.2011.00893.x.

Pyka, A. (2002) Innovation networks in economics: from the incentive-based to the knowledge based approaches. *European Journal of Innovation Management*, **5**, 3, 152–163.

Righi, C. and Simcoe, T. (2019) Patent examiner specialization. *Research Policy*, **48**, 1, 137–148. https://doi.org/10.1016/j.respol.2018.08.003.

Rosenkopf, L. and Almeida, P. (2003) Overcoming local search through alliances and mobility. *Management Science*, **49**, 6, 751–766. https://doi.org/10.1287/mnsc.49.6.751.16026.

Stewart, T.A. (1997) *Intellectual Capital: The New Wealth of Organizations*. New York, NY: Doubleday-Currency.

Stuart, T.E. and Podolny, J.M. (1996) Local search and the evolution of technological capabilities. *Strategic Management Journal*, **17**, S1, 21–38. https://doi.org/10.1002/smj.4250171004.

USPTO. (2018) *Ex Parte Reexamination Communication Transmittal Form*. https://www.documentcloud.org/documents/4936884-2018-Sep-12-Final-Rejection.html

USPTO. (2019) *Patent Process Overview*. https://www.uspto.gov/patents-getting-started/patent-process-overview{\#}step7

USPTO. (2020) *Unexamined Patent Application Inventory*. https://www.uspto.gov/corda/dashboards/patents/main.dashxml?CTNAVID=1005

USPTO. (2023) *Responding to Office Actions*. https://www.uspto.gov/patents/maintain/responding-office-actions

USPTO Press Release. (2007) *USPTO to Test Impact of Public Input on Improving Patent Quality in the Computer Technologies*. http://www.uspto.gov/news/pr/2007/07-21.jsp

Wired. (2017) *The Spectator Who Threw a Wrench in the Waymo/Uber Lawsuit*. https://www.wired.com/story/eric-swildens-uber-waymo-lawsuit-patent/

Yamauchi, I. and Nagaoka, S. (2015) Does the outsourcing of prior art search increase the efficiency of patent examination? Evidence from Japan. *Research Policy*, **44**, 1601–1614. https://doi.org/10.1016/j.respol.2015.05.003.

Notes

1 Prior art does not necessarily exist physically. It is enough that someone, sometime, somewhere previously has described or shown.

2 "An office action is written correspondence from the patent examiner that requires a properly signed written response from the applicant in order for prosecution of the application to continue. Moreover, the reply must be responsive to each ground of rejection and objection made by the examiner." (USPTO, 2023)

[3] Following previous research, patent families were taken into consideration by deleting patents with the same abstract as a previously filed patent, which resulted in 207,365 observations (including 327 granted PTP patents) (Kaplan and Vakili, 2015). Unfortunately, abstracts were not available for the first dataset – using PublicPAIR – so the detection of patent families based on abstracts was not possible for dataset 1.

4 To enable a robust topic model and a more fine-grained content analysis, I defined 1,000 topics and followed previous research to clean the aforementioned abstracts (Kaplan and Vakili, 2015).

Hannes W. Lampe is a researcher at the Institute of Entrepreneurship at the Hamburg University of Technology. He received his doctorate in 2015 from the University of Hamburg. His research interests include interdisciplinary topics between Innovation Management as well as Entrepreneurship on the one hand, and Public Management on the one hand. His work has been published in the *Journal of Business Economics*, the *Journal of Small Business Management*, the *European Journal of Operational Research*, *International Journal of Public Sector Performance Management*, among others.

