Learning by Evaluating:
Evidence from Academic Peer Review

Please do not circulate without authors permission

Abstract

The evaluation of innovative projects is an essential task for scientific and business organizations alike. While prior work has focused on the quality of evaluations and how they affect innovators, little is known about how the evaluation process affects evaluators themselves. We propose that evaluations provide evaluators with valuable learning opportunities by exposing them to relevant, cutting-edge knowledge. We test the argument in the setting of academic peer review, using administrative data from the Institute of Physics Publishing comprising 104,306 reviewer-manuscript pairs across 55 physical sciences journals. We find that evaluating a manuscript more than doubles the likelihood of a reviewer using that knowledge (by citation) within three years compared to qualified evaluators who were invited but unavailable. Evaluators' geographic and intellectual distance to the work mediates the learning effects of evaluation. While unavailable evaluators from the same country as a paper's corresponding author have significantly higher baseline citation rates, evaluators from different countries who reviewed the paper achieved similarly high citation levels, indicating that structured evaluation of external ideas can compensate for geographic boundaries in knowledge transfer. Intellectual proximity between evaluators' recent work and the work being evaluated is associated with stronger learning effects. These findings show that expert evaluations benefit not only organizations (by improving resource allocation) but also the evaluators themselves, which helps explain why they participate in the laborious and seemingly under-incentivized work.

Keywords: Peer review, knowledge spillovers, geographic proximity, scientific collaboration, absorptive capacity

1. Introduction

Allocating resources is a core task for individuals and organizations. Whether in funding agencies, business firms, research institutions, or government bodies, the ability to identify and support promising opportunities from a pool of competing alternatives directly shapes an organization's innovation trajectory, competitive position, and long-term viability (Stephan, 2012; Criscuolo et al., 2017; Lane et al., 2022). In science, this allocation takes various forms—from determining which manuscripts deserve limited journal space and associated reputational benefits to which research proposals warrant scarce funding. The efficiency and effectiveness of these allocation decisions fundamentally influence the pace and direction of scientific progress in our knowledge economy (Myers, 2020; Ayoubi et al., 2020).

To accomplish this critical function, organizations increasingly rely on evaluation systems that leverage expert judgment to assess the quality, potential, and fit of proposals, projects, or manuscripts (Boudreau et al., 2016; Erosheva et al., 2021). From hiring decisions and R&D project selection to venture capital investments and scientific peer review, expert evaluation has become a ubiquitous mechanism underpinning diverse organizational processes. The implicit assumption is that expert evaluators provide unique insights that improve allocation decisions—justifying the substantial resources dedicated to these assessment systems (Li, 2017).

Yet these evaluation processes create an interesting paradox: while they benefit organizations through improved resource allocation, the review process often goes unrecognized and is viewed primarily as a burden by the experts themselves (Bendiscioli, 2019). The mounting demand for experts' judgment is particularly alarming given growing concerns about "reviewer fatigue" in scientific publishing (Kovanis et al., 2016; Vesper, 2018). Peer review imposes considerable demands on evaluators' time and cognitive resources—resources that early-career researchers could otherwise devote to publications, grants, and other activities directly rewarded in tenure and promotion decisions (Franzoni, Scellato, and Stephan, 2011; Stephan, 1996; Walton, 2019). Given these pressures, many organizations are exploring ways to reduce reliance on expert evaluators, with some even considering automated alternatives like artificial

intelligence (Lane et al., 2024; Doshi et al., 2024). While AI-based solutions offer promising efficiency gains, they also introduce substantial uncertainties around quality and reliability of evaluations. More importantly, the rush toward automation may overlook a critical dimension of the peer review process: its potential role as a knowledge acquisition channel with valuable spillover benefits for reviewers. Survey evidence suggests that only 6% of reviewers believe the process enhances their professional credentials, with most citing intrinsic motivations (Walton, 2019). However, the persistence of reviewer participation despite minimal formal incentives hints at unrecognized benefits beyond relational contracting, editor relationships, and altruism (Engers and Gans, 1998; Griffin and Barczak, 2020; Squazzoni et al., 2013)—benefits that may include learning opportunities.

We address this puzzle by investigating peer review as a knowledge acquisition channel with significant evaluation knowledge spillovers. Our approach conceptualizes peer review as a structured commitment device that "forces" reviewers to engage deeply with new material (Bryan, Karlan, and Nelson, 2010), similar to how conferences facilitate focused engagement (Teplitskiy et al., 2023). This commitment likely yields substantial learning benefits that extend beyond altruistic or reputational incentives, potentially addressing a key gap in our understanding of knowledge flow mechanisms in scientific communities (Murray & O'Mahony, 2007). Our study addresses three central questions: (1) Does reviewing a manuscript increase the likelihood of knowledge acquisition, as evidenced by subsequent citations? (2) How does geographic distance moderate this relationship—does reviewing compensate for or reinforce existing knowledge flows across spatial boundaries? (3) Does the intellectual distance between a reviewer's expertise and the manuscript's content influence the magnitude of learning?

With these questions we posit a knowledge acquisition through peer review that is not uniformly distributed across all evaluator-evaluated relationships. The efficacy of this learning channel likely depends on two critical dimensions of distance between reviewer and manuscript. First, intellectual distance—the degree of overlap between the reviewer's expertise and the manuscript's subject matter—may influence knowledge acquisition dynamics (Leahey, Beckman, and Stanko, 2017). Greater distance potentially exposes reviewers to non-redundant knowledge that stimulates creativity (Jeppesen and

Lakhani, 2010), while minimal distance may limit novel learning opportunities. That said, we recognize that greater intellectual distance creates a tension: while it may expose reviewers to more novel content, it likely produces higher variance in scientific utility and provides fewer opportunities for reviewers to deploy this knowledge in their own work. Rather than assuming distant knowledge is inherently more useful, we expect it may have a higher-tail but lower-mean distribution of usefulness (Fleming, 2001).

Second, geographic distance may shape knowledge transfer during review in ways that complement or substitute for other knowledge diffusion mechanisms. Reviewers from different regions bring diverse contextual experiences that could enhance their understanding and integration of new ideas (Catalini et al., 2020). Geographic separation often correlates with differences in institutional cultures, methodological approaches, and research priorities—all of which might enrich the reviewer's learning experience (Breschi and Lissoni, 2009). These distance effects may be particularly important in determining how knowledge flows through structured evaluation processes across organizational and geographic boundaries (Lane & Maxfield, 1996).

To empirically test these relationships, we leverage administrative data from 60 physical sciences journals published by the Institute of Physics and employ an innovative identification strategy. Our analysis compares reviewers who completed evaluations with qualified individuals who were invited but unavailable—providing a natural experiment to isolate the causal effect of reviewing on knowledge acquisition. This approach reveals that reviewing a manuscript more than doubles the likelihood of citing it within three years—strong evidence of knowledge acquisition. Moreover, both geographic and intellectual proximity moderate this effect in revealing ways. While reviewers from the same country as the manuscript's authors show higher baseline citation rates, the act of reviewing partially compensates for barriers imposed by geographic distance—the treated "far" group (distant reviewers who evaluated the manuscript) achieves citation levels similar to the control "close" group (proximate reviewers who did not evaluate). Similarly, intellectual proximity (measured through abstract similarity) demonstrates a positive baseline effect, with reviewing amplifying this relationship.

These findings contribute to several important research streams with implications for both theory and practice. First and foremost, they reveal a previously underappreciated private benefit of evaluating – learning—which helps explain the paradox of continued participation in seemingly under-incentivized review activities. This informs the literature on contributions to public goods by identifying a concrete non-altruistic benefit that helps address concerns about reviewer fatigue. It also enriches our understanding of evaluation processes in innovation by highlighting benefits beyond resource allocation. Second, the effects vary with geographic and intellectual distance, informing the literature on geography and knowledge diffusion by demonstrating how formal evaluation processes can partly overcome spatial barriers to knowledge transfer. It also contributes to research on absorptive capacity by showing how intellectual proximity enhances learning benefits, supporting the importance of prior related knowledge in effectively assimilating and applying new information.

The remainder of the paper proceeds as follows. Section 2 develops our theoretical framework and formal hypotheses. Section 3 describes our research design and empirical approach. Section 4 details the variables and provides descriptive statistics. Section 5 presents our findings, and Section 6 concludes with discussion and implications for theory and practice.

2. Theoretical Framework and Hypotheses

2.1 Knowledge Acquisition Through Peer Review

Prior work suggests that scientific discovery is a process that may arise through multiple pathways, such as interactions with other researchers, experimentation, reading published work, applying to grants, or attending seminars and conferences (Boudreau and Lakhani, 2015; Dahlander and McFarland, 2013; Ayoubi et al., 2019; Lane et al., 2020; Teplitskiy et al., 2023). Consequently, it suggests that peer evaluation is an alternative pathway through which individuals are exposed to new ideas that can be absorbed, learned, and integrated into their own research, and may result in knowledge transfer.

A key insight in investigating when encounters with novel ideas will affect innovative potential is acknowledging that exposure is only half the story; the other half requires integrating and applying new

knowledge with one's own experiences and expertise (Argote, 2012; Argote and Ingram, 2000; Bresman, 2010). It is only after both of these stages have been carried out effectively that learning has occurred (Argote, 2012; Argote and Ingram, 2000; Huber, 1991). This process is often evidenced by knowledge recombination (Fleming, 2001; Schilling and Green, 2011; Weitzman, 1998). Knowledge recombination involves the movement of facts, relationships, ideas and insights from one setting to another, and becomes evident when the experience acquired by an individual, group or firm in one setting is *integrated* into another setting (Argote, 2012; Hansen, 1999; Szulanski, 1996). Much research has conceptualized knowledge transfer and integration as a two-sided process that depends on the source's willingness to share knowledge with a recipient and the recipient's efforts to acquire, absorb and integrate it with their existing knowledge bases (Argote, 2012). Hence, the process of knowledge transfer tends to be difficult to facilitate effectively – and suggests that even when individuals are exposed to diverse information and new ideas, they may fail to acquire the knowledge and integrate it effectively with their existing knowledge base (Carlile, 2004; Kogut and Zander, 1992).

There are two channels through which exposures to new ideas can lead to knowledge transfer. The first channel is through personal interactions or direct contact, such as face-to-face, phone or video conference (Hansen, 1999; Reagans and McEvily, 2003, 2003; Uzzi, 1997). These personal interactions occur when one person shares advice or information with another individual and they are often informal in nature (Haas and Hansen, 2007). Although these types of interactions are important for developing social relations between people (Uzzi, 1997; Uzzi and Spiro, 2005), one drawback is that because the messages are in real-time and simultaneous, the communication is fleeting and cannot be reviewed and revisited (Clark and Brennan, 1991). Hence, these types of interactions may not lead to knowledge transfer because the ideas, insights and advice exchanged are typically not recorded and cannot be reviewed after the initial encounter and exchange.

The second channel is through exposure to the ideas contained in written documents (Haas and Hansen, 2007), and is the focus of our study. Written documents play an essential role in knowledge sharing within organizations, and are an important way through which individuals are exposed to

diverse information and ideas. Knowledge workers often disseminate what they know in writing, such as in presentations, policies, tutorials, and upload these documents into databases or document management systems that can then be accessed by third parties (Haas and Hansen, 2007). A key advantage of written documents is that the communication is persistent and remains accessible in the same form as the original display for others to search, browse, annotate, visualize, structure and recontextualize (Erickson and Kellogg, 2000). Whereas verbal communication fades quickly, written documents stay behind as an artifact that can be reviewed later by any individual (Clark and Brennan, 1991). For example, Jeff Bezos, the long-time Amazon CEO, has famously instilled a culture of written documents before important meetings. In particular, rather than brainstorming ideas, Amazon executives are asked to submit six-page, dense, narratively structured memos of their novel ideas that each senior leader reads prior to starting a discussion (Bashaw, 2019). While both verbal and written communication channels have their merits, what's most critical for knowledge transfer in evaluation contexts is the depth of engagement with novel ideas. Evaluation processes are particularly powerful because they require evaluators to move beyond passive consumption of content to active critical assessment. When evaluating manuscripts, reviewers must thoroughly understand the work's arguments, methods, and implications, analyze its contributions relative to existing knowledge, and form reasoned judgments about its validity and importance. This structured, deliberate engagement creates cognitive conditions that are especially conducive to knowledge absorption and integration, regardless of the specific medium through which the ideas are conveyed. In academic peer review specifically, the responsibility of providing constructive feedback further incentivizes deep engagement with the material, potentially enhancing learning outcomes beyond what might occur through casual reading or conversation. This engaged, critical interaction with novel ideas helps establish common knowledge and beliefs between the evaluator and the knowledge creator, and has been shown to improve the transmission of complex ideas (Clark and Brennan, 1991), such as new scientific concepts. These reasons suggest that evaluations of novel ideas, embodied in scientific articles, are likely to facilitate exposure to diverse ideas that leads to successful knowledge transfer outcomes.

The persistent nature of communication in written documents helps to establish common knowledge and beliefs among two parties, and has been shown to improve the transmission of complex ideas (Clark and Brennan, 1991), such as new scientific concepts. These reasons suggest that evaluations of novel ideas, embodied in scientific articles, are likely to facilitate exposure to diverse ideas that leads to successful knowledge transfer outcomes.

In academic science, written documents are an important means through which knowledge is disseminated, as evidenced by the vast volume of scientific material published each year (Stephan, 2012). Integral to this process is the role of evaluations in scrutinizing the knowledge claims of others in the field, and is understood as providing "a system of institutionalized vigilance" where evaluators provide an impartial and fair interpretation of the submitted work (Merton, 1973). There are two milestones in the progression of a scientific idea that typically involves formal evaluations by experts. The first is for prospective evaluations of early-stage ideas, which entails decisions of whether to support them for funding and further development (Ayoubi, Pezzoni, and Visentin, 2019; Boudreau *et al.*, 2016; Li, 2017). The second is for submitted manuscripts which undergo retrospective evaluations of completed manuscripts prior to publication (Blank, 1991; Cicchetti, 1991). We argue that prospective evaluations of novel scientific manuscripts submitted to journals create unique opportunities for exposure to diverse ideas.

The knowledge embodied in scientific articles, namely novel ideas, is by definition, comprised of new research that aims to extend the current state of knowledge beyond the current knowledge frontier (Boudreau *et al.*, 2016). This suggests that the novel nature of the ideas contained in papers makes them well-suited for idea exposures that may eventually lead to successful knowledge transfer. In particular, studies on creativity and innovation suggest that people's initial ideas are often not the most innovative ones, because people tend to default to prior experiences when coming up with their earliest ideas. Hence, it is only after these preliminary ideas have been exhausted that more promising ideas emerge (Ward, 1994). This creates opportunities for evaluators to not only absorb and acquire new ideas, but also reuse, reshape, refine and combine them into their own work. Therefore, we hypothesize:

Hypothesis 1 (H1): Reviewers acquire new knowledge through the evaluation of submissions, as evidenced by an increased likelihood of citing reviewed papers in their subsequent publications.

As firms and individuals search for new knowledge, both local searches that concentrate narrower domains and distant searches spanning more diverse domains can lead to encounters with novel ideas. Hence, beyond the role of exposure to new ideas on knowledge acquisition, it is essential to understand how encounters with novel ideas in both local and distant searches for new knowledge affect knowledge acquisition – to gain deeper insight into the innovation process. In other words, it is critical to consider the impact of the geographical as well as the intellectual distance between the firm or individual's domain of expertise and the focal domain of the idea on knowledge acquisition.

In the remainder of this section, we theorize when exposure to diverse information sources will lead to knowledge transfer. More specifically, we theorize how two factors shape whether encounters with novel ideas are likely to result in knowledge recombination: the national proximity of the producer and evaluator, and the intellectual distance between an individual's expertise and the domain of the article evaluated.

2.2 The Moderating Role of Geographic Distance

Having established that peer review can serve as a mechanism for knowledge acquisition in Section 2.1, we now examine how the effectiveness of this channel varies with geographic distance between reviewers and manuscript authors. While the structured engagement of peer review may facilitate knowledge transfer broadly, the magnitude of this effect likely depends on pre-existing knowledge access pathways that differ systematically by geography. Geographic proximity has long been recognized as a catalyst for knowledge spillovers across organizations, institutions, and individual researchers (Jaffe, Trajtenberg, & Henderson, 1993; Audretsch & Feldman, 1996; Singh, 2005). Unlike the formal, structured knowledge transfer that occurs during peer review, proximity enables numerous informal pathways—frequent face-to-face interactions, casual encounters, and participation in local knowledge networks—that facilitate the spontaneous transmission of both codified and tacit knowledge (Maskell & Malmberg,

1999; Storper & Venables, 2004; Howells, 2002). These informal, geography-based knowledge flows operate through multiple complementary channels. Co-location reduces coordination costs and travel time, increasing the frequency of in-person exchanges (Catalini, 2018; Rawley & Seamans, 2015) that facilitate the transmission of complex, tacit knowledge resistant to codification in written documents (Roche, 2020). Shared regional contexts foster common understanding and interpretive frameworks that ease knowledge assimilation (Saxenian, 1996; Gertler, 2003), while local professional networks create redundant information pathways that increase serendipitous exposure to relevant work (Powell, Koput, & Smith-Doerr, 1996; Owen-Smith & Powell, 2004). These mechanisms collectively ensure that scientists are typically more aware of—and learn from—research conducted in their immediate geographic vicinity, independent of formal review processes (Breschi & Lissoni, 2009; Almeida & Kogut, 1999). Recent research provides compelling causal evidence for these geographic effects. Catalini, Fons-Rosen, and Gaulé (2020) leveraged quasi-random variation in scientific conference timing to demonstrate that even temporary colocation significantly increases subsequent citations between previously unconnected scientists. This effect persists even in today's digitally connected world, underscoring geography's enduring role in knowledge diffusion. Similarly, Roche (2020) demonstrated that geographic proximity uniquely benefits knowledge recombination—inventors relocating to innovation hubs experienced significant increases in combinatorial novelty by accessing locally embedded knowledge that facilitated unexpected connections between previously separate domains. Even within single institutions, the microgeography of physical space shapes knowledge flows. Catalini (2018) showed that relocating scientists to different floors within the same building dramatically reduced their collaboration likelihood, while moves to adjacent spaces increased collaborative activities. These findings indicate that informal knowledge spillovers remain highly localized even within institutional boundaries—a phenomenon with direct implications for understanding how geographic proximity shapes baseline knowledge access among potential reviewers. For geographically proximate reviewers and authors, these overlapping informal channels create significant redundancy in knowledge access pathways. Proximate reviewers may encounter authors' work through local seminars, institutional repositories, pre-publication workshops, or mutual

collaborators (Boudreau et al., 2017; Ganguli, 2017) long before formal review processes begin. Shared cultural contexts and institutional norms further reduce barriers to knowledge absorption (Phelps, Heidl, & Wadhwa, 2012). As a result, reviewers from the same geographic region as manuscript authors likely possess higher baseline familiarity with the research content before any formal review begins. In stark contrast, geographically distant reviewers typically lack these abundant informal exposure channels. Without the structured engagement provided by peer review, they might never encounter—or deeply engage with—research from distant locales (Bell & Zaheer, 2007; Cummings & Kiesler, 2007). For these distant reviewers, the formal review process serves as a crucial bridge across geographical divides, providing access to knowledge that might otherwise remain entirely inaccessible (Agrawal, Kapur, & McHale, 2008; Fleming, King, & Juda, 2007). While geographically proximate reviewers can draw on multiple redundant information sources, distant reviewers often have the peer review process as their primary or sole channel for exposure to specific research developments from other regions (Kotha, George, & Srikanth, 2013; Tortoriello, Reagans, & McEvily, 2012). This fundamental disparity in baseline knowledge access suggests a key interaction effect between the review process and geographic distance. While reviewing likely increases knowledge acquisition for all evaluators compared to non-reviewing, the marginal information gain should be systematically larger for geographically distant reviewers who lack alternative exposure pathways (Broekel & Boschma, 2012; Laursen, Reichstein, & Salter, 2011). In essence, peer review may serve as a partial substitute for the knowledge access advantages that geographic proximity typically provides (Roche, 2020; Catalini et al., 2020), with distant reviewers having "more to learn" from the structured engagement with manuscripts outside their geographic reach. HERE!! Make this flow better to H2. Conversely, local reviewers—already embedded in overlapping knowledge networks with authors—may experience comparatively smaller learning increments from the formal review process (Boschma, 2005; Torre & Rallet, 2005). For these proximate reviewers, the manuscript content may contain fewer genuinely novel elements, as they may have already absorbed key ideas through their rich local information environment. Thus, while they still benefit from the structured engagement that peer review provides (as argued in Hypothesis 1), the magnitude of this benefit should be relatively smaller compared to that experienced by distant reviewers. Recent empirical

evidence supports this differential effect across various structured knowledge exchange contexts. Boudreau et al. (2017) documented how structured research symposia produced greater collaboration benefits for previously unconnected researchers. Roche, Oettl, and Catalini (2024) further demonstrated that temporary colocation at technology coworking hubs yields disproportionate knowledge transfer benefits for scientists working in less connected institutions. These findings consistently show that formal knowledge exchange mechanisms yield their greatest marginal impact when they bridge pre-existing geographic divides. Research by Catalini et al. (2020) on travel costs and scientific collaboration further illuminates this relationship. Analyzing the introduction of low-cost airline routes between U.S. cities, they found that reduced travel costs significantly increased cross-institutional collaboration in affected city pairs, with the strongest effects for collaborations involving specialized knowledge that would be difficult to exchange through digital channels alone. This evidence underscores how structured interaction opportunities—whether through reduced travel costs or formal review processes—can compensate for the absence of spontaneous, proximity-based knowledge exposure. One can therefore expect that exposure to certain knowledge through reviewing will be reinforced by the closer geographical distance with the source of the knowledge. These effects should be even greater given other aspects of national proximity such as common institutional norms, scientific systems, language and social networks.

Building on these theoretical foundation and empirical evidence, we propose that the knowledge acquisition benefits of peer review (established in Hypothesis 1) will vary systematically with geographic distance. The "learning-by-reviewing" effect should be relatively stronger for reviewers who are geographically proximate to manuscript authors, as they start with higher baseline knowledge access through multiple informal channels. This leads to our second hypothesis:

Hypothesis 2 (H2). Geographic distance negatively moderates the learning-by-reviewing effect. Reviewers from a different country than the manuscript's authors experience a smaller "boost" in subsequent citations than do reviewers from the same country.

2.3 The Moderating Role of Intellectual Distance

Having established the role of geographic distance in moderating knowledge acquisition through peer review, we now turn to a second critical dimension: intellectual distance—the cognitive gap between a reviewer's expertise domain and a manuscript's content. While geographic proximity facilitates knowledge access through physical channels, intellectual proximity shapes the cognitive processes involved in recognizing, assimilating, and applying new knowledge from manuscripts.

When reviewers evaluate manuscripts closely aligned with their expertise, cognitive proximity enables rapid comprehension of specialized terminology, methodological approaches, and theoretical foundations (Lane & Lubatkin, 1998; Reagans & McEvily, 2003). This familiarity reduces absorption costs and facilitates integration of new insights into existing knowledge structures. Cohen and Levinthal's (1990) work on absorptive capacity emphasizes that prior related knowledge enhances the assimilation of new information, suggesting that domain experts could more efficiently integrate novel contributions from manuscripts.

However, cognitive proximity also creates potential limitations for learning. When reviewers possess knowledge bases highly similar to manuscript authors, they may encounter fewer genuinely novel elements worth adopting. Low cognitive distance entails similar knowledge stocks and thus the probability of learning if there is too little novelty to learn from (Ayoubi et al., 2017). This redundancy effect potentially diminishes the marginal knowledge gain from review activities (Wuyts et al., 2005; Van Wijk, Jansen, & Lyles, 2008).

Conversely, extensive research on creativity highlights the generative potential of cognitive distance. When reviewers encounter manuscripts from adjacent domains, they gain exposure to non-redundant knowledge that may complement their existing expertise in unexpected ways (Hargadon & Sutton, 1997; Fleming, 2001). This exposure can trigger novel associations and recombinations that might not occur within their primary field (Amabile et al., 2005; Perry-Smith & Mannucci, 2017). Studies of

breakthrough innovations consistently find that many originate at the intersection of distant knowledge domains (Powell, Koput, & Smith-Doerr, 1996; Uzzi et al., 2013).

Yet excessive cognitive distance introduces comprehension barriers. When the gap between reviewer expertise and manuscript content becomes too large, reviewers may struggle to understand fundamental concepts, potentially hampering effective knowledge transfer (Zhou & Li, 2012; Piezunka & Dahlander, 2015). Cognitive load theory suggests that learning efficiency decreases when individuals must simultaneously master numerous unfamiliar concepts (Sweller, 1988; Paas, Renkl, & Sweller, 2003).

These competing mechanisms—enhanced novelty versus diminished comprehension—create a theoretical tension characterized by two countervailing effects: the "speaking the same language" effect that enhances knowledge flows when cognitive distance is low, and the "opening new horizons" effect that has a positive impact as cognitive distance increases (Ayoubi et al., 2017). Similar patterns emerge across multiple innovation contexts. Rosenkopf and Nerkar (2001) found that patents combining moderate technological distance yielded greater impact than those at either extreme. Fleming (2001) documented parallel findings in inventor productivity, with moderate knowledge recombination distance producing the most valuable innovations. In scientific evaluation, Boudreau et al. (2016) observed that reviewers responded most positively to proposals of intermediate novelty.

The peer review process introduces a unique dynamic to this relationship. Unlike established knowledge domains, manuscripts under review represent unpublished frontier research containing novel elements not yet in the mainstream literature. This inherent novelty partially mitigates redundancy concerns present in other knowledge exchange contexts. Even when reviewers possess substantial domain expertise, manuscripts likely offer new findings, methods, or theoretical extensions worth adopting. This distinctive characteristic suggests that while extreme intellectual distance may still impede learning through comprehension barriers, the redundancy risks at low intellectual distances are somewhat attenuated in peer review.

Building on these theoretical foundations and empirical evidence, we propose that the knowledge acquisition benefits established in Hypothesis 1 will be moderated by intellectual distance between reviewer and manuscript, with intellectual distance having a positive interaction on learning given that intellectually similar reviewers possess sufficient domain knowledge to comprehend the manuscript while still encountering novel elements that stimulate innovative recombinations. Given these theoretical considerations and empirical precedents, we posit:

Hypothesis 3 (H3). Intellectual distance between a reviewer's prior expertise and the manuscript negatively moderates the learning-by-reviewing effect. Reviewers with a high intellectual similarity to the manuscript being evaluated experience stronger knowledge gains than do those with low intellectual similarity.

3. Methodology

Having established our theoretical framework on how evaluating affects knowledge acquisition and the moderating roles of geographic and intellectual distance, we now detail our empirical approach to testing these hypotheses. This section outlines our data collection process, variable construction, and analytical strategy.

3.1 Data Collection

Our study utilizes administrative data from 62 Institute of Physics Publishing (IOP) journals, covering 229,173 manuscript submissions between 2018-2022. We implemented a multi-stage sample selection process to construct a dataset suitable for analyzing learning dynamics in peer review (see Figure 1). Data from IOP was then matched to records in OpenAlex—a comprehensive, open index of scholarly works (Priem, Piwowar, & Orr, 2022)—to construct several key variables.

First, we restricted our sample to include only manuscripts that (1) underwent at least one round of peer review (e.g., desk rejections excluded), (2) were submitted during 2018-2019 to avoid potential COVID-19 confounding effects while ensuring sufficient post-publication citation windows, and (3) were eventually published, either in IOP journals or elsewhere. For IOP-published papers (71,123

submissions), we used the DOI provided in the dataset to locate them directly in the OpenAlex database. For papers rejected from IOP but published elsewhere (93,605 submissions), we employed a two-step identification process employed in prior work (Chen et al., 2024). To link these rejected submissions to OpenAlex records, the first step involved creating an inclusive group of candidate records from the publications of all authors on a submission. If authors provided an ORCID, their record was located directly, otherwise author full name and a combination of author last name with country and field information were used for search. The second step involved the use of a logistic classifier trained on survey data (Chen et al., 2024) where submitting authors reported any name changes to their rejected submission. This two-step process achieved 97% precision and 78% recall in identifying published papers among candidate publications.

Second, for each submission, we collected the OpenAlex records of all works that cited the published version as of January 2025 ("citing works"). Submissions that received no citations were removed from the sample. Crucially, this step gathered full author lists for each citing work.

""Third, to identify reviewers in the OpenAlex database, we followed a systematic approach. Some reviewers provided an ORCID, allowing for direct location. For reviewers without an ORCID in our data, we conducted a search using institution name plus reviewer full name when institution data was available and found in OpenAlex. If institution data was missing or not found in OpenAlex, we resorted for reviewer full name search only. Searches specified the Physics concept tag to restrict results to relevant fields. We retained the first search result when fewer than 25 potential matches were found and excluded cases with too many potential matches (labeled "Too many authors found") or no matches ("No OA author found"). This process allowed us to collect productivity metrics as of January 2025 for 69,627 unique reviewers across 388,311 reviewer-manuscript pairs.

Finally, we included only invited reviewers who either (1) agreed to the invitation and completed a review (treatment group) or (2) declined the invitation explicitly stating they were unavailable (control group). Other responses not included in the final analytic sample include those who agreed to the invitation but did not complete a review, those who did not respond to invitation or auto-declined, and

those who explicitly declined for other reasons (e.g., conflict of interest, out of field). This restriction is crucial for our identification strategy, as unavailable reviewers represent a plausible counterfactual—qualified individuals who would have reviewed the submission but could not due to exogenous time constraints. In sum, this approach yielded a sample of 104,306 reviewer-manuscript pairs for 45,435 unique submissions to 55 journals for our primary analysis.

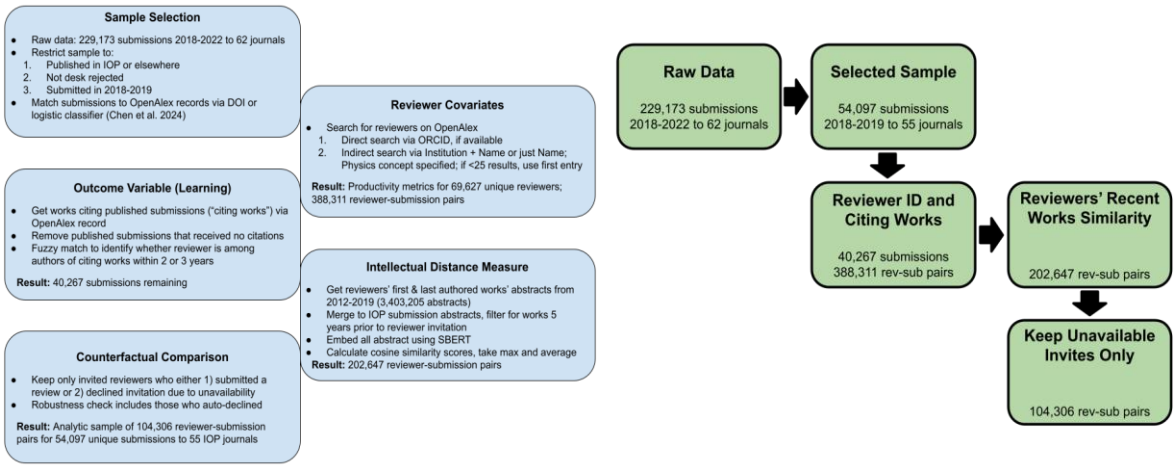


Figure 1: Data collection and cleaning process

3.2 Measurement of Variables

3.2.1 Dependent Variable: Learning

Our primary dependent variable is a binary indicator of reviewer learning, equal to 1 if the reviewer cited the manuscript they were invited to review in their subsequent work within three years, and 0 otherwise. This operationalization aligns with our theoretical framework, where citation behavior serves as a proxy for knowledge acquisition and integration.

To construct this variable, we restricted each submissions list of citing works to only keep those within three years of the reviewer's invitation. Then, reviewer names were fuzzy-matched to authors of the citing works using a Levenshtein distance threshold of 80. A similar process was used with a two-year window and to generate counts instead of binaries for later robustness checks.

3.2.2 Independent Variable: Review Completion

Our key independent variable is a binary indicator of review completion, derived from the Review.Invitation.Response field in our dataset (see Appendix for full list of responses). It equals 1 if the reviewer agreed to and completed the review, and 0 if they declined due to unavailability. More specifically, this variable is set to 1 if Review.Invitation.Response is "Accept" or "Agreed" AND the reviewer provided a recommendation on whether to accept or reject the submission. It is set to 0 if Review.Invitation.Response is "Declined – Unavailable." This variable allows us to test Hypothesis 1 on whether reviewing a manuscript increases the likelihood of subsequent knowledge acquisition.

3.2.3 Moderating Variables

Geographic Distance

To test Hypothesis 2 on the moderating role of geographic proximity, we constructed a binary variable indicating whether the reviewer and the manuscript's corresponding author are from the same country. We compared the country codes of the reviewer and the corresponding author, assigning a value of 1 when they matched, indicating geographic proximity, and 0 if they did not match. Most authors and reviewers in our data are from the USA, China, and India, but a total of 139 countries are represented in our analytic sample[1].

Intellectual Distance

To test Hypothesis 3 on the moderating role of intellectual proximity, we developed a sophisticated measure comparing the semantic similarity between submission abstracts and reviewers' recent publication abstracts. First, abstracts of submissions were already included in the IOP data. Reviewers' OpenAlex records were used to collect records of their first and last-authored publications from 2012-2019, totaling '3,403,205 abstracts for 69,627 unique reviewers.

[1] While acknowledging that same-country matching represents a coarse measure of geographic distance—researchers near borders may be physically closer to foreign colleagues than to distant compatriots—this approach still captures many proximity-related mechanisms relevant to knowledge flows, including shared institutional contexts, language, and research networks.

Second, for each reviewer-submission pair, the reviewer's abstracts were filtered to include only those from the five years prior to their invitation (e.g., 2013-2017 if reviewer was invited in 2018). This operationalizes a comparison targeting reviewers' recent work that is not influenced by reviewing.

Third, all abstracts (submission and reviewer's) are transformed into embeddings using the Sentence Transformers (SBERT) "all-mpnet-base-v2" model (Reimers & Gurevych, 2019). Embedding models fine-tuned for scientific text are often trained using citation networks and thus are less appropriate for our analysis of citation likelihood than more general models like SBERT.

Last, cosine similarity scores were calculated for each submission abstract and the reviewer's eligible publication abstracts, resulting in each reviewer-submission pair having an array of similarity scores. From this array, the maximum similarity score is used as our primary measure of intellectual proximity representing the minimum distance between the evaluator and the object of evaluation. Higher values indicate greater cognitive overlap. The average of the similarity score arrays were also captured for robustness checks. This procedure resulted in 202,647 reviewer-manuscript pairs with intellectual distance measures.

3.2.4 Control Variables

We included several control variables to account for factors that might influence citation patterns:

- *Team Size*: The number of authors on the manuscript, directly obtained from the data field indicating total authors.
- *Submission Year*: Extracted from the submission date information to account for temporal dynamics.
- *Journal Fixed Effects*: Controls for unobserved heterogeneity at the journal level.
- *Manuscript Fixed Effect*: Controls for unobserved heterogeneity at the manuscript level including topic and quality. Use of manuscript fixed effects absorbs other controls like team size and submission year, as these do not vary within-manuscript.

- *Reviewer Productivity Measures*: Including counts of publications and citations, h-index, and i-10-index. Metrics are binned into quintiles for deidentification purposes.[2]

3.3 Analytical Approach

We employ OLS regression models to estimate the probability of a reviewer citing a manuscript based on reviewing status and our moderating variables. Our baseline model for testing Hypothesis 1 takes the form:

$$P(Learning_i = 1) = (\beta_0 + \beta_1 \times Reviewed_i + \beta_2 \times TeamSize_i + \beta_3 \times SubmissionYear_i + Reviewer\ Productivity\ Measures + Fixed\ Effects + \varepsilon_i)$$

Where Learning_i is the binary learning indicator for reviewer i, Reviewed_i indicates whether reviewer i completed the review, and the remaining terms represent control variables and fixed effects. The coefficient β_1 measures the effect of reviewing on learning, with a positive and significant value supporting Hypothesis 1.

To test Hypothesis 2 on the moderating role of geographic proximity, we extend the model to include an interaction term:

$$P(Learning_i = 1) = (\alpha_0 + \alpha_1 \times Reviewed_i + \alpha_2 \times SameCountry_i + \alpha_3 \times (Reviewed_i \times SameCountry_i) + \alpha_4 \times TeamSize_i + \alpha_5 \times SubmissionYear_i + Reviewer\ Productivity\ Measures + Fixed\ Effects + \varepsilon_i)$$

The coefficient α_3 captures how geographic proximity moderates the effect of reviewing on learning. A negative and significant α_3 would support Hypothesis 2, indicating that geographic distance negatively moderates the learning-by-reviewing effect.

Similarly, for Hypothesis 3 on the moderating role of intellectual distance, we estimate:

[2] For the published dataset, we implemented a deidentification process that: 1) transformed journal abbreviation, manuscript ID, reviewer name, and country information into random 10-character strings and 2) binned identifying covariates (total citations, publications, h-index, team size) into 5 quantiles.

$$P(Learning_i = 1) = (\gamma_0 + \gamma_1 \times Reviewed_i + \gamma_2 \times IntellProximity_i + \gamma_3 \times (Reviewed_i \times IntellProximity_i) + \gamma_4 \times TeamSize_i + \gamma_5 \times SubmissionYear_i + Reviewer\ Productivity\ Measures + Fixed\ Effects + \varepsilon_i)$$

The coefficient γ_3 captures how intellectual proximity moderates the effect of reviewing on learning. A positive and significant γ_3 would support Hypothesis 3, indicating that reviewers with high intellectual similarity to the manuscript experience stronger knowledge gains.

To address potential concerns about selection bias and endogeneity, we conduct several robustness checks, including alternative specifications of our dependent variable (different citation windows, citation count instead of binary), different sample restrictions (e.g., including reviewers who auto-declined or who agreed but did not submit reviews) and additional control variables and fixed effects. These specifications allow us to rigorously test our hypotheses while accounting for potential confounding factors, providing a comprehensive assessment of the learning-by-reviewing relationship and its contingencies.

4. Results

4.1 Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Full Sample					
Reviewer Learned (3yr)	104,306	0.023	0.150	0	1
Submission Year	104,306	2018.561	0.496	2018	2019
Final Decision	104,306	0.565	0.496	0	1
Citations (decile)	104,306	2.015	1.383	0	4
Works (decile)	104,306	2.016	1.383	0	4
h-index (decile)	104,306	1.978	1.379	0	4
i10-index (decile)	104,306	2.011	1.380	0	4
Team Size (decile)	104,306	1.796	1.332	0	4
Same Country	104,306	0.112	0.315	0	1
Max Abstract Similarity	96,975	0.522	0.171	-0.165	0.943
Did Not Review Paper (Unavailable for review)					
Reviewer Learned (3yr)	44,896	0.013	0.114	0	1
Submission Year	44,896	2018.562	0.496	2018	2019
Final Decision	44,896	0.631	0.483	0	1
Citations (decile)	44,896	2.195	1.375	0	4
Works (decile)	44,896	2.137	1.368	0	4
h-index (decile)	44,896	2.137	1.378	0	4
i10-index (decile)	44,896	2.148	1.367	0	4
Team Size (decile)	44,896	1.805	1.334	0	4
Same Country	44,896	0.080	0.271	0	1
Max Abstract Similarity	41,725	0.522	0.168	-0.158	0.943
Reviewed Paper (Submitted review)					
Reviewer Learned (3yr)	59,410	0.031	0.173	0	1
Submission Year	59,410	2018.560	0.496	2018	2019
Final Decision	59,410	0.515	0.500	0	1
Citations (decile)	59,410	1.878	1.374	0	4
Works (decile)	59,410	1.924	1.387	0	4
h-index (decile)	59,410	1.859	1.368	0	4
i10-index (decile)	59,410	1.907	1.380	0	4
Team Size (decile)	59,410	1.790	1.330	0	4
Same Country	59,410	0.136	0.343	0	1
Max Abstract Similarity	55,250	0.523	0.173	-0.165	0.927

Table 1: Descriptive statistics of key variables in the analytic sample

Our analytic sample consists of 104,306 reviewer-manuscript pairs across 55 journals in the physical sciences. Before presenting our regression analyses, we examine the patterns of learning-by-reviewing in our data through descriptive statistics.

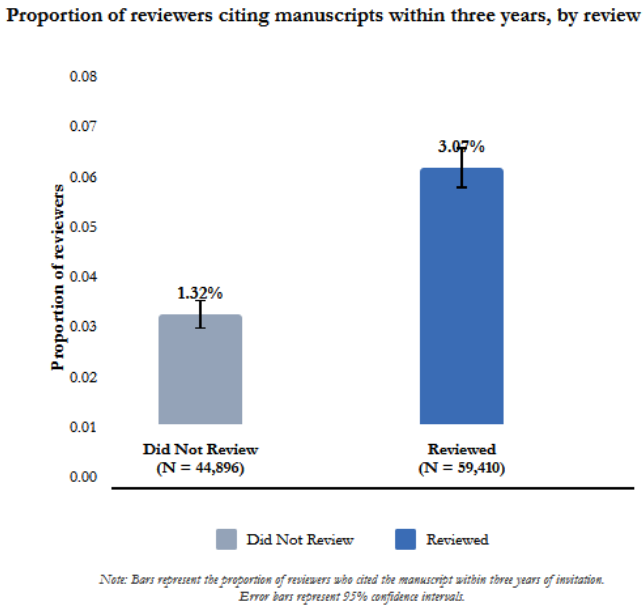
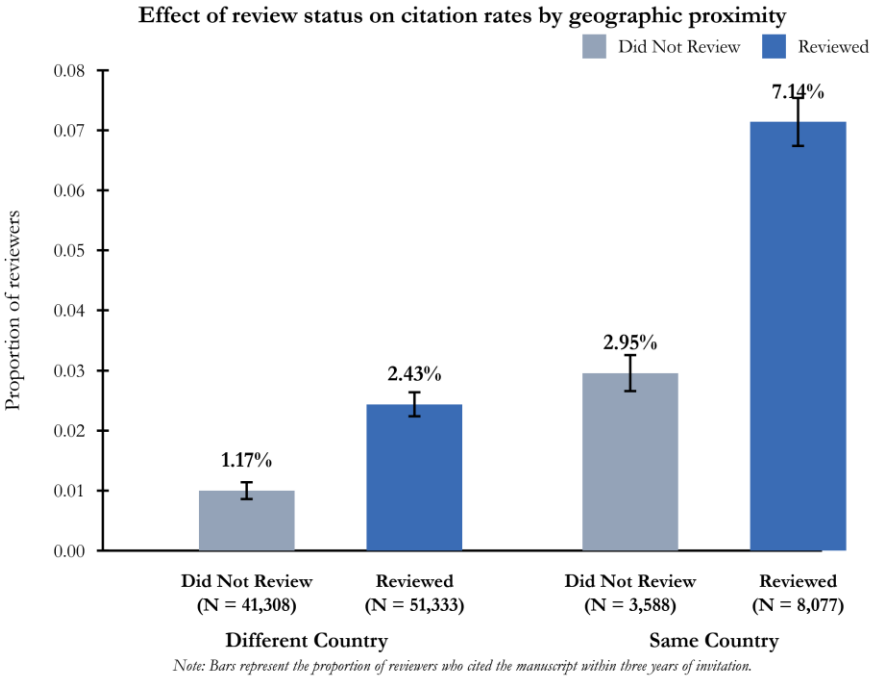


Figure 2 shows the proportion of reviewers who cite the manuscripts they were invited to review within three years, comparing those who completed reviews (treatment group) with those who declined due to unavailability (control group). The raw difference suggests that reviewers who evaluate manuscripts are more likely to subsequently cite them compared to qualified reviewers who were invited but unavailable.



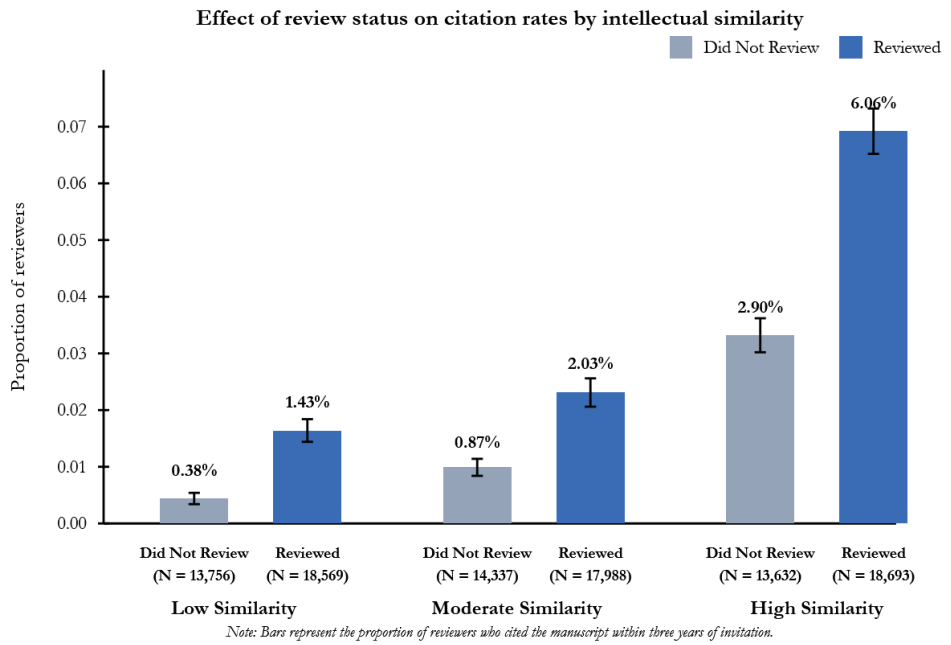


Figure 3a and 3b further breaks down citation rates by our key moderating variables: geographic proximity and intellectual distance. These descriptive patterns provide initial evidence for our hypothesized relationships, though multivariate analyses are necessary to account for potential confounding factors.

4.2 Main Results

4.2.1 Effect of Reviewing on Knowledge Acquisition

Variable	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Reviewed	0.018*** (0.001)	0.019*** (0.001)	0.020*** (0.001)	0.019*** (0.001)
Team Size		0.001*** (0.000)	0.001*** (0.000)	0.001*** (0.000)
Submission Year		-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)
Constant	0.013*** (0.001)	1.692 (1.896)	1.484 (1.897)	1.695 (1.896)
Controls		Yes	Yes	Yes
Journal FE			Yes	
Manuscript FE				Yes
Observations	104,306	104,306	104,306	104,306
Pseudo R^2	0.003	0.007	0.007	0.007

*Note: Standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05, † p<0.10*

Table 2: OLS regression results for Hypothesis 1

Table 2 presents the results of our OLS regression models testing Hypothesis 1, which predicted that reviewers acquire knowledge through manuscript evaluation. Model 1 includes only reviewer characteristics and journal fixed effects, while Model 2 adds manuscript-level controls.

Across both specifications, we find that completing a review significantly increases the likelihood of citing the manuscript in subsequent work. The coefficient on our Reviewed variable is positive and statistically significant ($p < 0.001$), indicating that, all else equal, reviewers who evaluate a manuscript are more likely to incorporate its knowledge compared to those who were invited but unavailable. This effect remains robust when controlling for reviewer productivity, manuscript characteristics, and journal fixed effects.

In terms of magnitude, the average marginal effect indicates that reviewing a manuscript increases the probability of subsequent citation by approximately 1.8 percentage points, representing more than a twofold increase relative to the baseline citation rate among non-reviewers. This substantial effect supports Hypothesis 1 and suggests that the peer review process serves as an important channel for knowledge acquisition.

4.2.2 Moderating Effect of Geographic Distance

Variable	Model 1	Model 2	Model 3	Model 4
Reviewed	0.013*** (0.001)	0.014*** (0.001)	0.015*** (0.001)	0.014*** (0.001)
Same Country	0.018*** (0.003)	0.017*** (0.003)	0.016*** (0.003)	0.017*** (0.003)
Reviewed × Same Country	0.029*** (0.003)	0.030*** (0.003)	0.030*** (0.003)	0.030*** (0.003)
Team Size		0.001*** (0.000)	0.001*** (0.000)	0.001*** (0.000)
Submission Year		-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)
Constant	0.012*** (0.001)	2.555 (1.890)	2.345 (1.891)	2.557 (1.890)
Controls		Yes	Yes	Yes
Journal FE			Yes	
Manuscript FE				Yes
Observations	104,306	104,306	104,306	104,306
Pseudo R^2	0.010	0.014	0.014	0.014

*Note: Standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05, † p<0.10*

Table 3: OLS regression results for Hypothesis 2

Table 3 presents models testing Hypothesis 2, which predicted that geographic distance negatively moderates the learning-by-reviewing effect. The models include the interaction term between *Reviewed* and *Same Country*, which captures how the effect of reviewing differs based on geographic proximity between reviewer and authors.

The coefficient on the interaction term is negative and statistically significant across specifications, indicating that the learning benefit of reviewing is smaller for reviewers from the same country as the manuscript authors compared to reviewers from different countries. This finding supports Hypothesis 2 and suggests that reviewing partially compensates for the barriers to knowledge flow imposed by geographic distance.

4.2.3 Moderating Effect of Intellectual Distance

Variable	Model 1	Model 2	Model 3	Model 4
Reviewed	-0.009** (0.003)	-0.009** (0.003)	-0.009** (0.003)	-0.009** (0.003)
Abstract Similarity	0.062*** (0.004)	0.056*** (0.004)	0.055*** (0.004)	0.056*** (0.004)
Reviewed × Abstract Similarity	0.052*** (0.006)	0.055*** (0.006)	0.055*** (0.006)	0.055*** (0.006)
Team Size		0.001** (0.000)	0.001** (0.000)	0.001** (0.000)
Submission Year		0.001 (0.001)	0.001 (0.001)	0.001 (0.001)
Constant	-0.019*** (0.002)	-1.606 (1.994)	-1.740 (1.994)	-1.603 (1.994)
Controls		Yes	Yes	Yes
Journal FE			Yes	
Manuscript FE				Yes
Observations	96,975	96,975	96,975	96,975
Pseudo R^2	0.015	0.017	0.018	0.017

*Note: Standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05, † p<0.10*

Table 4: OLS regression results for Hypothesis 3

Table 4 presents models testing Hypothesis 3, which predicted that intellectual proximity positively moderates the learning-by-reviewing effect. These models include the interaction between *Reviewed* and *Intellectual Proximity*, our measure of cognitive overlap between reviewer expertise and manuscript content.

The coefficient on this interaction term is positive and statistically significant, indicating that reviewers with higher intellectual similarity to the manuscript content experience stronger knowledge gains from reviewing. This result supports Hypothesis 3 and suggests that while reviewing generally facilitates learning, the effect is amplified when reviewers possess relevant domain knowledge that enables them to effectively assimilate new information.

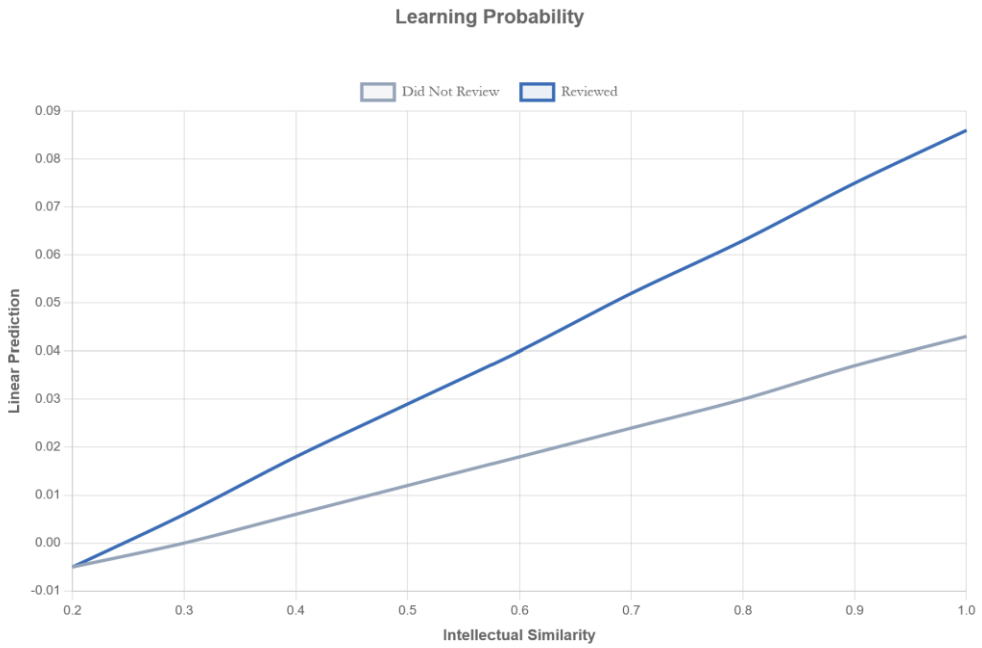


Figure 4: Marginal effects of reviewing on citation probability by intellectual proximity

Figure 4 visualizes how the effect of reviewing varies across different levels of intellectual proximity. The steeper slope at higher levels of intellectual proximity demonstrates that reviewers with greater cognitive overlap with the manuscript content derive greater learning benefits from the review process.

4.3 Robustness Checks

To ensure the validity of our findings, we conducted several robustness checks. First, we tested alternative time windows for measuring citations (2-year and unlimited windows) and found consistent results. Second, we ran logistic specification instead of OLS and our main findings remained robust. Third, we used a count variable for the number of times the paper reviewed is cited rather than the binary variable, again finding support for our hypotheses. Finally we expanded our sample to include reviewers who declined for reasons other than unavailability with robust findings.

Additionally, we tested for potential selection bias using Heckman selection models and found no evidence that our results are driven by systematic differences between reviewers who accept versus

decline invitations. These robustness checks strengthen our confidence in the causal interpretation of our findings.

5. Discussion

This study advances our understanding of expert evaluation by highlighting its role not only as a resource allocation mechanism but also as a significant channel for knowledge transfer. Our results demonstrate that reviewers acquire knowledge through the evaluation process, as evidenced by their increased likelihood of citing reviewed manuscripts in subsequent work. Moreover, we find that both geographic and intellectual distance moderate this learning effect in important ways.

Our findings contribute to several streams of literature at the intersection of knowledge management, scientific collaboration, and peer review processes. First, we extend research on knowledge transfer mechanisms by identifying peer review as a structured process that facilitates deep engagement with new ideas. While existing work has focused on channels such as conferences, collaborations, and mobility (Lane et al., 2020; Catalini, 2018), the peer review process represents a unique and widespread mechanism that has received limited attention as a knowledge transfer channel.

Second, our findings on the moderating role of geographic distance contribute to the literature on spatial constraints in knowledge flows. Previous research has documented how geographic proximity facilitates knowledge transfer through increased interaction opportunities and shared contextual understanding (Breschi & Lissoni, 2009; Catalini, Fons-Rosen, & Gaulé, 2020). Our results suggest that formal review processes can partially compensate for these geographic barriers, providing a structured channel for knowledge to flow across distant locations. Interestingly, we find that while unavailable evaluators from the same country as a paper's corresponding author have significantly higher baseline citation rates, evaluators from different countries who reviewed the paper achieved similarly high citation levels. This indicates that structured evaluation of external ideas can compensate for geographic boundaries in knowledge transfer. The effect also appears larger for proximate reviewers, likely because "same-country" does not necessarily mean close proximity but does confer advantages like shared time

zones, language, culture, and conference participation opportunities. This insight is particularly relevant as science becomes increasingly global, with researchers collaborating across institutional and national boundaries.

Third, our results on intellectual proximity contribute to debates about optimal cognitive distance in knowledge exchange. While some scholars emphasize the benefits of diverse knowledge bases for innovation (Jeppesen & Lakhani, 2010; Fleming, 2001), others highlight the importance of shared understanding for effective knowledge transfer (Cohen & Levinthal, 1990; Lane & Lubatkin, 1998). Our finding that intellectual proximity enhances learning-by-evaluating suggests that in the context of peer review—where manuscripts contain inherently novel elements—the ability to comprehend and assimilate new information may outweigh the potential benefits of extreme cognitive distance.

Beyond these theoretical contributions, our findings connect to several influential streams of research. First, they build on absorptive capacity theory, which emphasizes the role of prior knowledge in assimilating new information (Cohen and Levinthal, 1990). Second, they relate to knowledge recombination perspectives, which highlight how innovation emerges from connecting previously separate ideas (Fleming, 2001; Uzzi, Mukherjee, Stringer, and Jones, 2013). Third, they engage with studies on distance-related barriers to knowledge sharing (Breschi and Lissoni, 2009; Catalini et al., 2020). Additionally, our work contributes to research on public goods provision and organizational learning. The literature on public goods has emphasized both altruistic and social motives for contribution, as seen in studies of Wikipedia (Zhang and Zhu, 2011), as well as private benefits like signaling in open-source software development (Lerner and Tirole, 2002). However, these findings have remained disconnected from evaluation contexts until our study. By identifying knowledge spillovers in evaluation processes, we provide a novel perspective on why individuals participate in seemingly under-incentivized activities (Engers and Gans, 1998).

These knowledge spillovers emerge as the peer review process operates as both a quality control mechanism and a potential learning opportunity for evaluators. Beyond its gatekeeping role in ensuring methodological rigor (Tomkins, Zhang, and Heavlin, 2017; Bol, de Vaan, and van de Rijt, 2018), peer

review serves as a conduit for knowledge acquisition. Reviewers engaging with cutting-edge research encounter novel ideas, methods, and perspectives that could stimulate their intellectual development and spark knowledge spillovers (Catalini et al., 2020; Cohen and Levinthal, 1990). This knowledge transfer potential aligns with organizational learning principles: just as firms learn through external interactions (Cohen and Levinthal, 1990; Argote, 2012; Hwang, Singh & Argote, 2015), reviewers can benefit from engaging with manuscripts spanning diverse fields, regions, and communities. The structured nature of peer review, requiring close attention to content, methods, and arguments, creates conditions conducive to deep learning (Bryan, Karlan, and Nelson, 2010)—potentially fostering creativity and new research directions through knowledge recombination (Fleming, 2001; Uzzi et al., 2013).

Our findings have important practical implications well beyond addressing "reviewer fatigue"—a growing concern in scientific publishing (Kovanis et al., 2016; Vesper, 2018). While editorial teams increasingly struggle to recruit qualified reviewers as demands on researchers' time intensify, and this is certainly an important challenge, our findings have broader relevance. Many organizations across diverse contexts conduct project evaluations using internal or external experts, making it crucial to understand the full set of costs and benefits associated with these evaluation systems. The learning benefits we document offer a compelling alternative narrative about the value of evaluation from the evaluator's perspective. Instead of viewing evaluation solely as a professional service obligation driven by altruism or reputation concerns, organizations and individuals might recognize it as an opportunity for knowledge acquisition and professional development.

Journal editors and publishers could leverage these insights in several ways. First, they might explicitly communicate the potential learning benefits when inviting reviewers, possibly increasing acceptance rates. Second, our findings on moderating factors suggest possibilities for strategic reviewer assignment: editors could consider geographic and intellectual proximity when matching reviewers to manuscripts, potentially optimizing both evaluation quality and knowledge dissemination. Third, implementing

reviewer training programs that emphasize effective knowledge extraction techniques could enhance both review quality and learning outcomes.

From a policy perspective, our findings suggest that evaluation systems serve dual purposes in scientific advancement. Beyond their gatekeeping function, they represent an important mechanism for knowledge dissemination across geographic and disciplinary boundaries. This insight is particularly important given recent trends toward delegating evaluations to artificial intelligence systems (Lane et al., 2024; Doshi et al., 2024). While AI-based solutions offer promising efficiency gains, the rush toward automation may overlook the critical role of peer review as a knowledge transfer channel with valuable spillover benefits for human reviewers. Funding agencies and institutions might consider more explicit recognition of reviewing activities in evaluation metrics, acknowledging their contribution not only to quality assurance but also to knowledge flow within scientific communities.

Despite these contributions, our study has several limitations that suggest directions for future research. First, while citation behavior provides an observable proxy for knowledge acquisition, it may not capture all forms of learning from the review process. Future studies might employ surveys, interviews, or text analysis methods to investigate more subtle forms of influence. Second, our analysis focuses on physics journals, and the patterns we observe may differ across scientific fields with varying epistemic cultures and review practices. Comparative studies across disciplines would enrich our understanding of how contextual factors shape learning-by-reviewing effects.

Additionally, our approach faces methodological limitations. We only observe evaluation at IOP journals and do not know if drafts were evaluated at other venues before or after IOP prior to publication. Furthermore, intellectual distance is proxied by semantic similarity; other operationalizations could include reference similarity, keyword similarity, co-author network distance, and other measures. Our analytical strategy using unavailable reviewers as a control group mitigates some selection concerns, but unobserved variables might still affect both reviewing likelihood and subsequent citation behavior. Future research using experimental approaches or natural experiments could provide additional causal evidence for the learning-by-reviewing effect.

6. Conclusion

Building on our theoretical contributions, this study provides compelling evidence that peer review serves as a significant knowledge acquisition channel—an insight with direct implications for addressing growing concerns about reviewer recruitment and retention in scientific publishing. Rather than viewing evaluation solely as an under-incentivized service, it can be recognized as a valuable opportunity for professional development and knowledge acquisition, with reviewers more than twice as likely to integrate evaluated knowledge into their own work. For policy makers, our findings support the explicit recognition of reviewing activities in academic evaluation metrics. The knowledge spillovers we document demonstrate that evaluation systems simultaneously serve resource allocation and knowledge dissemination functions, providing empirical justification for policies that incentivize participation in peer review. The moderating effects of geographic and intellectual proximity further suggest opportunities for strategic reviewer assignment to optimize both evaluation quality and knowledge flow.

Beyond addressing "reviewer fatigue," these insights extend to organizational settings broadly, offering a novel perspective on why individuals participate in seemingly under-incentivized activities. As scientific and business organizations increasingly explore automated alternatives to human evaluation, our study highlights important but often overlooked benefits of traditional expert review systems. Future research should examine how these learning dynamics manifest across disciplines and institutional contexts, and how evaluation systems might be intentionally designed to enhance knowledge spillovers without compromising assessment integrity. By documenting the multifaceted value of peer review, this study contributes to ongoing efforts to sustain effective evaluation systems amid mounting demands on experts' time and attention.

References

Agrawal, A., Kapur, D., & McHale, J. (2008). How do spatial and social proximity influence knowledge flows? Evidence from patent data. *Journal of Urban Economics, 64*(2), 258-269.

Almeida P, Kogut B. 1999. Localization of knowledge and the mobility of engineers in regional networks. *Management Science 45*(7): 905–917.

Amabile, T. M., Barsade, S. G., Mueller, J. S., & Staw, B. M. (2005). Affect and creativity at work. *Administrative science quarterly, 50*(3), 367-403.

Argote L. 2012. Organizational learning: Creating, retaining and transferring knowledge. Springer Science & Business Media.

Argote L, Ingram P. 2000. Knowledge transfer: A basis for competitive advantage in firms. *Organizational Behavior and Human Decision Processes 82*(1): 150–169.

Audretsch DB, Feldman MP. 1996. R&D spillovers and the geography of innovation and production. *American Economic Review 86*(3): 630–640.

Ayoubi C, Pezzoni M, Visentin F. 2017. At the origins of learning: Absorbing knowledge flows from within the team. *Journal of Economic Behavior & Organization 134:* 374–387.

Ayoubi C, Pezzoni M, Visentin F. 2019. The important thing is not to win, it is to take part: What if scientists benefit from participating in research grant competitions? *Research Policy 48*(1): 84–97.

Ayoubi, C., Barbosu, S., Pezzoni, M., & Visentin, F. (2020). *What matters in funding: The value of research coherence and alignment in evaluators' decisions.* (Mentioned as "(Ayoubi et al., 2020)" in the text.)

Bell GG, Zaheer A. 2007. Geography, networks, and knowledge flow. *Organization Science 18*(6): 955–972.

Bendiscioli S. 2019. The troubles with peer review for allocating research funding: Funders need to experiment with versions of peer review and decision-making. *EMBO Reports 20*(12): e49472.

Blank RM. 1991. The effects of double-blind versus single-blind reviewing: Experimental evidence from the American Economic Review. *The American Economic Review 81*(5): 1041–1067.

Bol T, de Vaan M, van de Rijt A. 2018. The Matthew effect in science funding. *Proceedings of the National Academy of Sciences 115*(19): 4887–4890.

Boschma R. 2005. Proximity and innovation: A critical assessment. *Regional Studies 39*(1): 61–74.

Boudreau KJ, Guinan EC, Lakhani KR, Riedl C. 2016. Looking across and looking beyond the knowledge frontier: Intellectual distance, novelty, and resource allocation in science. *Management*

Science 62(10): 2765–2783.

Boudreau KJ, Lakhani KR. 2015. "Open" disclosure of innovations, incentives and follow-on reuse: Theory on processes of cumulative innovation and a field experiment in computational biology. *Research Policy 44*(1): 4–19.

Boudreau, K. J., Brady, T., Ganguli, I., Gaule, P., Guinan, E., Hollenberg, A., & Lakhani, K. R. (2017). A field experiment on search costs and the formation of scientific collaborations. *Review of Economics and Statistics, 99*(4), 565-576.

Breschi S, Lissoni F. 2009. Mobility of skilled workers and co-invention networks: An anatomy of localized knowledge flows. *Journal of Economic Geography 9*(4): 439–468.

Bresman H. 2010. External learning activities and team performance: A multimethod field study. *Organization Science 21*(1): 81–96.

Broekel T, Boschma R. 2012. Knowledge networks in the Dutch aviation industry: The proximity paradox. *Journal of Economic Geography 12*(2): 409–433.

Carlile PR. 2004. Transferring, translating, and transforming: An integrative framework for managing knowledge across boundaries. *Organization Science 15*(5): 555–568.

Catalini C. 2018. Microgeography and the direction of inventive activity. *Management Science 64*(9): 4348–4364.

Catalini C, Fons-Rosen C, Gaulé P. 2020. How do travel costs shape collaboration? *Management Science 66*(8): 3340–3360.

Chen, H., Rider, C. I., Jurgens, D., & Teplitskiy, M. (2024). Geographical disparities in navigating rejection in science drive disparities in its file drawer. Available at SSRN.

Cicchetti DV. 1991. The reliability of peer review for manuscript and grant submissions: A cross-disciplinary investigation. *Behavioral and Brain Sciences 14*(1): 119–135.

Clark HH, Brennan SE. 1991. Grounding in communication. In *Perspectives on socially shared cognition* (Resnick LB, Levine JM, Teasley SD, eds.). American Psychological Association, pp. 127–149.

Cohen WM, Levinthal DA. 1990. Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly 35*(1): 128–152.

Criscuolo P, Dahlander L, Grohsjean T, Salter A. 2017. Evaluating novelty: The role of panels in the selection of R&D projects. *Academy of Management Journal 60*(2): 433–460.

Cummings JN, Kiesler S. 2007. Coordination costs and project outcomes in multi-university collaborations. *Research Policy 36*(10): 1620–1634.

Dahlander L, McFarland DA. 2013. Ties that last: Tie formation and persistence in research collaborations over time. *Administrative Science Quarterly 58*(1): 69–110.

Doshi AR, Bell JJ, Mirzayev E, Vanneste BS. 2025. Generative artificial intelligence and evaluating strategic decisions. *Strategic Management Journal 46*(3): 583–610.

Engers M, Gans JS. 1998. Why referees are not paid (enough). *The American Economic Review 88*(5): 1341–1349.

Erickson T, Kellogg WA. 2000. Social translucence: An approach to designing systems that support social processes. *ACM Transactions on Computer-Human Interaction (TOCHI) 7*(1): 59–83.

Erosheva, E. A., Grant, S., Chen, M. C., Lindner, M. D., Nakamura, R. K., & Lee, C. J. (2020). NIH peer review: Criterion scores completely account for racial disparities in overall impact scores. *Science Advances, 6*(23), eaaz4868.

Fleming L. 2001. Recombinant uncertainty in technological search. *Management Science 47*(1): 117–132.

Fleming L, King C, Juda AI. 2007. Small worlds and regional innovation. *Organization Science 18*(6): 938–954.

Franzoni C, Scellato G, Stephan P. 2011. Changing incentives to publish. *Science 333*(6043): 702–703.

Ganguli I. 2017. Immigration and ideas: What did Russian scientists "bring" to the United States? *Journal of Labor Economics 35*(S1): S299–S341.

Gertler MS. 2003. Tacit knowledge and the economic geography of context. *Journal of Economic Geography 3*(1): 75–99.

Griffin A, Barczak G. 2020. Effective reviewing for conceptual journal submissions. *AMS Review 10*(1): 36–48.

Haas MR, Hansen MT. 2007. Different knowledge, different benefits: Toward a productivity perspective on knowledge sharing in organizations. *Strategic Management Journal 28*(11): 1133–1153.

Hansen MT. 1999. The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits. *Administrative Science Quarterly 44*(1): 82–111.

Hargadon A, Sutton RI. 1997. Technology brokering and innovation in a product development firm. *Administrative Science Quarterly 42*(4): 716–749.

Howells J. 2002. Tacit knowledge, innovation and economic geography. *Urban Studies 39*(5–6): 871–884.

Huber GP. 1991. Organizational learning: The contributing processes and the literatures. *Organization*

Science 2(1): 88–115.

Jeppesen LB, Lakhani KR. 2010. Marginality and problem-solving effectiveness in broadcast search. *Organization Science 21*(5): 1016–1033.

Jaffe AB, Trajtenberg M, Henderson R. 1993. Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly Journal of Economics 108*(3): 577–598.

Kogut B, Zander U. 1992. Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science 3*(3): 383–397.

Kotha R, George G, Srikanth K. 2013. Bridging the mutual knowledge gap: Coordination and the commercialization of radical science in new ventures. *Organization Science 24*(5): 1291–1310.

Kovanis M, Porcher R, Ravaud P, Trinquart L. 2016. The global burden of journal peer review in the biomedical literature: strong imbalance in the collective enterprise. *PloS ONE 11*(11): e0166387.

Laursen K, Reichstein T, Salter A. 2011. Exploring the effect of geographical proximity and knowledge search strategies on innovative performance. *Regional Studies 45*(1): 33–49.

Lane JN, Boussioux L, Ayoubi C, Chen YH, Len C, Spens R, Wang PH, et al. 2024. The Explainability Edge? A Field Experiment on Generative AI-Augmented Evaluations of Early-Stage Innovations. *Harvard Business School Working Paper (25-001).*

Lane JN, Ganguli I, Gaule P, Guinan E, Lakhani KR. 2020. Engineering serendipity: When does knowledge sharing lead to knowledge production? *Strategic Management Journal 41*(3): 531–560.

Lane, J. N., Teplitskiy, M., Gray, G., Ranu, H., Menietti, M., Guinan, E. C., & Lakhani, K. R. (2022). Conservatism gets funded? A field experiment on the role of negative information in novel project evaluation. *Management science, 68*(6), 4478-4495.

Lane PJ, Lubatkin M. 1998. Relative absorptive capacity and interorganizational learning. *Strategic Management Journal 19*(5): 461–477.

Leahey E, Beckman CM, Stanko TL. 2017. Prominent but less productive: The impact of interdisciplinarity on scientists' research. *Administrative Science Quarterly 62*(1): 105–139.

Lerner J, Tirole J. 2002. Some simple economics of open source. *Journal of Industrial Economics 50*(2): 197–234.

Li D. 2017. Expertise versus bias in evaluation: Evidence from the NIH. *American Economic Journal: Applied Economics 9*(2): 60–92.

Maskell P, Malmberg A. 1999. The competitiveness of firms and regions: 'Ubiquitification' and the importance of localized learning. *European Urban and Regional Studies 6*(1): 9–25.

Merton RK. 1973. The sociology of science: Theoretical and empirical investigations. University of Chicago Press.

Mikolov T, Chen K, Corrado G, Dean J. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Myers K. 2020. The elasticity of science. *American Economic Journal: Applied Economics 12*(4): 103–134.

Owen-Smith J, Powell WW. 2004. Knowledge networks as channels and conduits: The effects of spillovers in the Boston biotechnology community. *Organization Science 15*(1): 5–21.

Paas F, Renkl A, Sweller J. 2003. Cognitive load theory and instructional design: Recent developments. *Educational Psychologist 38*(1): 1–4.

Perry-Smith JE, Mannucci PV. 2017. From creativity to innovation: The social network drivers of the four phases of the idea journey. *Academy of Management Review 42*(1): 53–79.

Piezunka H, Dahlander L. 2015. Distant search, narrow attention: How crowding alters organizations' filtering of suggestions in crowdsourcing. *Academy of Management Journal 58*(3): 856–880.

Powell WW, Koput KW, Smith-Doerr L. 1996. Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology. *Administrative Science Quarterly 41*(1): 116–145.

Priem J, Piwowar H, Orr R. 2022. OpenAlex: A fully open index of scholarly works, authors, and venues. *arXiv preprint arXiv:2205.01833*.

Rawley E, Seamans R. 2015. Curtailing curtailment: An empirical analysis of the sunk-cost effect. *Management Science 61*(8): 1922–1939.

Reagans R, McEvily B. 2003. Network structure and knowledge transfer: The effects of cohesion and range. *Administrative Science Quarterly 48*(2): 240–267.

Reimers N, Gurevych I. 2019. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. *arXiv preprint arXiv:1908.10084*.

Roche MP. 2020. Taking innovation to the streets: Microgeography, physical structure, and innovation. *Review of Economics and Statistics 102*(5): 912–928.

Roche, M. P., Oettl, A., & Catalini, C. (2024). Proximate (co-) working: Knowledge spillovers and social interactions. *Management Science, 70*(12), 8245-8264.

Rosenkopf L, Nerkar A. 2001. Beyond local search: Boundary-spanning, exploration, and impact in the optical disk industry. *Strategic Management Journal 22*(4): 287–306.

Saxenian A. 1996. Regional advantage: Culture and competition in Silicon Valley and Route 128.

Harvard University Press.

Schilling MA, Green E. 2011. Recombinant search and breakthrough idea generation: An analysis of high impact papers in the social sciences. *Research Policy 40*(10): 1321–1331.

Singh J. 2005. Collaborative networks as determinants of knowledge diffusion patterns. *Management Science 51*(5): 756–770.

Squazzoni F, Bravo G, Takács K. 2013. Does incentive provision increase the quality of peer review? An experimental study. *Research Policy 42*(1): 287–294.

Stephan PE. 1996. The economics of science. *Journal of Economic Literature 34*(3): 1199–1235.

Stephan PE. 2012. How economics shapes science. Harvard University Press, Cambridge, MA.

Storper M, Venables AJ. 2004. Buzz: Face-to-face contact and the urban economy. *Journal of Economic Geography 4*(4): 351–370.

Szulanski G. 1996. Exploring internal stickiness: Impediments to the transfer of best practice within the firm. *Strategic Management Journal 17*(S2): 27–43.

Teplitskiy M, Park S, Thompson N, Karger D. 2023. The effect of conference presentations on the diffusion of ideas. Working Paper.

Tomkins A, Zhang M, Heavlin WD. 2017. Reviewer bias in single-versus double-blind peer review. *Proceedings of the National Academy of Sciences 114*(48): 12708–12713.

Torre A, Rallet A. 2005. Proximity and localization. *Regional Studies 39*(1): 47–59.

Tortoriello M, Reagans R, McEvily B. 2012. Bridging the knowledge gap: The influence of strong ties, network cohesion, and network range on the transfer of knowledge between organizational units. *Organization Science 23*(4): 1024–1039.

Uzzi B. 1997. Social structure and competition in interfirm networks: The paradox of embeddedness. *Administrative Science Quarterly 42*(1): 35–67.

Uzzi B, Mukherjee S, Stringer M, Jones B. 2013. Atypical combinations and scientific impact. *Science 342*(6157): 468–472.

Uzzi B, Spiro J. 2005. Collaboration and creativity: The small world problem. *American Journal of Sociology 111*(2): 447–504.

Van Wijk R, Jansen JJ, Lyles MA. 2008. Inter- and intra-organizational knowledge transfer: A meta-analytic review and assessment of its antecedents and consequences. *Journal of Management Studies 45*(4): 830–853.

Vesper I. 2018, September 7. Peer reviewers unmasked: Largest global survey reveals trends. *Nature.*

Available at: https://www.nature.com/articles/d41586-018-06602-y.

Walton R. 2019, April 16. Recognising the contribution of Nature Research journal referees. *Nature Research Blog.* http://blogs.nature.com/ofschemesandmemes/2019/04/16/recognising-the-contribution-of-nature-research-journal-referees

Ward TB. 1994. Structured imagination: The role of category structure in exemplar generation. *Cognitive Psychology 27*(1): 1–40.

Weitzman ML. 1998. Recombinant growth. *The Quarterly Journal of Economics 113*(2): 331–360.

Wuyts S, Dutta S, Stremersch S. 2005. Portfolios of interfirm agreements in technology-intensive markets: Consequences for innovation and profitability. *Journal of Marketing 69*(2): 46–59.

Zhang X, Zhu F. 2011. Group size and incentives to contribute: A natural experiment at Chinese Wikipedia. *American Economic Review 101*(4): 1601–1615.

Zhou KZ, Li C. 2012. How knowledge is transferred in R&D alliances: A multi-relational social capital perspective. *Journal of Product Innovation Management 29*(4): 591–607.