



Strategy Science

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

Organizing Knowledge Production Teams Within Firms for Innovation

Vikas A. Aggarwal, David H. Hsu, Andy Wu

To cite this article:

Vikas A. Aggarwal, David H. Hsu, Andy Wu (2020) Organizing Knowledge Production Teams Within Firms for Innovation. *Strategy Science* 5(1):1-16. <https://doi.org/10.1287/stsc.2019.0095>

This work is licensed under a Creative Commons Attribution 4.0 International License. You are free to copy, distribute, transmit and adapt this work, but you must attribute this work as “*Strategy Science*. Copyright © 2020 The Author(s). <https://doi.org/10.1287/stsc.2019.0095>, used under a Creative Commons Attribution License: <https://creativecommons.org/licenses/by/4.0/>.”

Copyright © 2020 The Author(s)

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes. For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

Organizing Knowledge Production Teams Within Firms for Innovation

Vikas A. Aggarwal,^a David H. Hsu,^b Andy Wu^c

^a INSEAD, 77305 Fontainebleau, France; ^b The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104;

^c Harvard Business School, Harvard University, Boston, Massachusetts 02163

Contact: vikas.aggarwal@insead.edu,  <https://orcid.org/0000-0002-2040-6767> (VAA); dhsu@wharton.upenn.edu,
 <https://orcid.org/0000-0001-8228-5449> (DHH); awu@hbs.edu,  <https://orcid.org/0000-0002-9107-5731> (AW)

Received: October 26, 2018

Revised: May 23, 2019; September 2, 2019

Accepted: September 6, 2019

Published Online in Articles in Advance:
March 6, 2020

<https://doi.org/10.1287/stsc.2019.0095>

Copyright: © 2020 The Author(s)

Abstract. How should firms organize their pool of inventive human capital for firm-level innovation? Although access to diverse knowledge may aid knowledge recombination, which can facilitate innovation, prior literature has focused primarily on one way of achieving that: diversity of inventor-held knowledge within a given knowledge production team (*within-team knowledge diversity*). We introduce the concept of *across-team knowledge diversity*, which captures the distribution of inventor knowledge diversity across production teams, an overlooked dimension of a firm's internal organization design. We study two contrasting forms of organizing the firm-level knowledge diversity environment in which a firm's inventors are situated: *diffuse* (high within-team diversity and low across-team diversity) versus *concentrated* (low within-team diversity and high across-team diversity). Using panel data on new biotechnology ventures founded over a 21-year period and followed annually from inception, we find that concentrated structures are associated with higher firm-level innovation quality, and with more equal contributions from their teams (and the opposite for diffuse structures). Our empirical tests of the operative mechanisms point to the importance of within-team coordination costs in diffuse structures and across-team knowledge flows in concentrated knowledge structures. We end with a discussion of implications for future research on organizing for innovation.



Open Access Statement: This work is licensed under a Creative Commons Attribution 4.0 International License. You are free to copy, distribute, transmit and adapt this work, but you must attribute this work as "Strategy Science. Copyright © 2020 The Author(s). <https://doi.org/10.1287/stsc.2019.0095>, used under a Creative Commons Attribution License: <https://creativecommons.org/licenses/by/4.0/>."

Funding: Funding from the Harvard Business School Division of Research and Faculty Development, the INSEAD-Wharton Center for Global Research and Education, the INSEAD Alumni Fund, and the Mack Institute for Innovation Management at the University of Pennsylvania, is gratefully acknowledged.

Keywords: knowledge recombination • organization design • teams • diversity • innovation • human capital

1. Introduction

Collaborative (rather than solo) knowledge production is increasingly the norm in many creative domains (Wuchty et al. 2007). Patenting follows this same trend, driven in part by the increasingly high cumulative inventor knowledge necessary to master fields (Jones 2009). With inventors becoming narrower but deeper in their knowledge domains, a team-based patent production structure may compensate. Empirical evidence suggests that teams are more likely than solo inventors to produce breakthrough inventions (Singh and Fleming 2010).

Invention studies in this science of science tradition (e.g., Fortunato et al. 2018) often use the population of U.S. inventors, and so the findings are quite general. However, these studies typically do not account for the organizational context in which inventive teams reside and are therefore silent on the managerially

relevant question of how production team organization within a firm relates to firm-level innovation outcomes. The literature on product development teams in technology-based environments (the "teams" literature) is typically much richer in the depth of data collection, often via interviews and surveys (e.g., Ancona and Caldwell 1992, Hoegl et al. 2004), but it faces drawbacks in generalizing the findings, especially across organizations. Moreover, in this branch of the literature, the usual unit of analysis is the team rather than the firm, limiting inferences regarding the innovation role of the organizational knowledge environments in which teams are embedded.

We aim to bridge this gap by layering an organizational and team design lens onto a relatively large empirical sample of invention teams and firms. Although the literature recognizes the central role of knowledge recombination in generating impactful

innovation, the typical instrument of accessing novel knowledge to recombine (Fleming 2001) is within-team knowledge diversity. By broadening the lens to the firm, we introduce a new concept to the firm-level organizing for innovation literature of sourcing knowledge for potential recombination: across-team knowledge diversity, which captures the extent to which invention teams differ from one another with respect to their prior technical domain experience. At the extreme, firms achieve knowledge diversity in two contrasting ways: (1) complete within-team diversity paired with no across-team diversity (what we term a *diffuse* organizational knowledge environment) or (2) no within-team diversity paired with complete across-team diversity (a *concentrated* environment). We analyze such organizational team structures and their association with innovation performance and distribution at the firm level.

Empirically, we draw on a sample of venture capital-backed biotechnology firms observed longitudinally from their date of founding, collecting information on the full career experience of these firms' inventors to measure individual-level knowledge. We proxy members of a production team through patent coinventor records. Using firm and year fixed effects specifications, we find that organizing teams in a diffuse manner is associated with lower forward patent citations (with more unequal contributions from individual patent teams toward those citations). We find the opposite pattern for concentrated firm knowledge environments. Consistent with our theorized mechanisms, we find that production teams in diffuse (versus concentrated) environments are more inert (less likely to reference inventions from other teams within the firm). Lowering within-team coordination costs (such as via prior collaborative experience) is thus associated with improved innovation performance. On the other hand, in concentrated organizational knowledge environments, the operative mechanism influencing innovation performance is across-team knowledge flows. Greater geographic span among a firm's inventors (which makes such flows more difficult) thus links to reduced innovation performance. In a concluding section, we discuss the implications of our findings for work on intrafirm networks and the knowledge-based view in the broader strategy literature.

2. Literature and Hypotheses

Although the literature on organizing for innovation has made substantial progress over the past several decades in developing a deep understanding of how external-to-the-firm factors such as alliances, acquisitions, and human capital mobility shape firm-level innovation, progress in understanding how a firm's internal organization shapes firm-level innovation

has been more fragmented. Studies in this vein have contributed important insights on factors such as R&D centralization or decentralization (e.g., Argyres and Silverman 2004, Leiponen and Helfat 2011), intrafirm network configuration (e.g., Tsai 2002), and top management team composition (e.g., Bantel and Jackson 1989). Yet data availability has likely constrained theorizing with regard to how production team organization within firms relates to firm-level innovation outcomes. We discuss two topics in this section: how knowledge diversity and access jointly create distinct organizational knowledge environments in which teams pursue innovation activities; and the link between these environments and firm-level innovation outcomes.

2.1. Organizational Knowledge Environments

We contend that the diversity of and access to knowledge within the firm must be considered jointly in order to understand the link between inventive human capital organization and innovation. First, the organization of human capital within the firm determines where the inventive process occurs. Second, given the locus of the inventive process, organization also determines the available knowledge upon which the team's inventive process operates. Team boundaries affect the motivation for and difficulty of accessing knowledge from others within the firm but outside the focal invention team.

Inventors embody knowledge accumulated through their past experiences (Gruber et al. 2013), with the aggregate set of inventive human capital inside a firm reflecting the available knowledge inventors can draw upon as part of the production process (Simon 1991, Grant 1996).¹ Under this view, the set of input knowledge to an invention is distributed throughout the firm. In this sense, a production team can build on knowledge that is local (contained within the team itself) or boundary-spanning (existing inside the firm but on a different production team). From a team design perspective, knowledge diversity both within a focal team and outside the focal team but within the firm, together with knowledge access in those same locales, will likely shape innovation output.² The access issues we discuss later relate to team boundaries, together with the relative ease and incentives to consult colleagues across versus within those team boundaries.

Prior literature suggests that the diversity of knowledge among a firm's inventors benefits knowledge production (Hoisl et al. 2016). Access to diverse sets of knowledge facilitates innovation because it expands the scope of available material for recombinant search (Schumpeter 1934, Hargadon and Sutton 1997, Carnabuci and Operti 2013). Accordingly, the technical experience diversity of a firm's inventors critically determines a firm's innovative capacity (Dixon 1994). Yet this line of research is agnostic to whether

intrafirm boundaries enable or impede the flow of knowledge available for recombination.

By contrast, another research stream suggests that the location of knowledge relative to an intrafirm boundary influences a firm’s innovation output by influencing the ease of search for this knowledge. For example, hierarchical structures can inhibit knowledge sharing (Tsai 2002), whereas centralized R&D functions allow inventors to draw on a wider range of technologies (Argyres and Silverman 2004). However, this literature on intraorganizational boundaries does not explicitly consider the role of knowledge diversity among the firm’s inventors. Locus of knowledge among teams within a firm immediately surfaces issues of access, with regard to both inventors’ motivation to access and to inventors’ ability to access and use knowledge outside the focal team. Knowledge that crosses team boundaries faces a higher hurdle than knowledge present within a team (Tushman 1977).

To understand how the organization of inventors inside a firm influences firm-level innovative output, we characterize the diversity of knowledge within a firm along two distinct dimensions. First, *within-team knowledge diversity* is the degree to which an average production team inside the firm has individuals who differ among themselves with respect to the knowledge embodied in their prior technical experience. Second, *across-team knowledge diversity* is the degree to which an average production team inside the firm differs from other teams with respect to the knowledge embodied in the aggregate technical experience of the team’s inventors. These forms of knowledge diversity can be conceptualized as characteristics of the firm-level environment in which production teams are embedded. In the former case, individuals

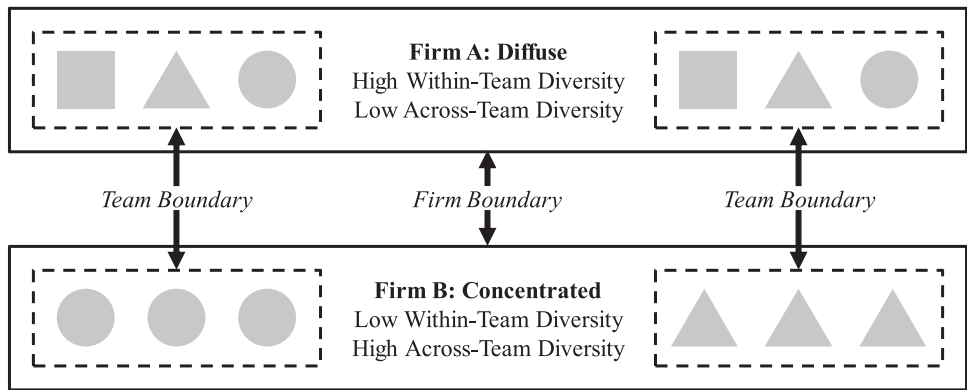
encounter knowledge diversity among their own team members. In the latter case, knowledge diversity resides across team boundaries. To ensure these concepts are meaningful when considered together, we confine our conceptual development and empirical analyses to firms organized into multiple production teams (Sabbagh 1996, Gerwin and Moffat 1997).

These dimensions of knowledge diversity allow us to characterize firms with regard to their distinctive organizational knowledge environments, which we propose shape firm-level innovation outcomes. Figure 1 depicts two contrasting organizational knowledge environments. The top panel illustrates a firm with high levels of within-team knowledge diversity and low levels of across-team knowledge diversity. This is a knowledge environment in which knowledge in a given technological area can be seen as diffuse throughout the firm. By contrast, the bottom panel illustrates a firm with low levels of within-team knowledge diversity and high levels of across-team knowledge diversity. This is a knowledge environment where knowledge can be characterized as concentrated, in the sense that teams hold unique knowledge relative to other teams in the firm. We turn next to the implications of these two distinct organizational knowledge environments.

2.2. Diffuse Organizational Knowledge Environments and Firm-Level Innovation

In diffuse organizational knowledge environments, the diversity of prior technical experience among inventors on a given production team means that these inventors may lack shared points of understanding and common building blocks that can help guide the inventive process. Such a structure may impede innovation quality due to within-team coordination challenges resulting from heterogeneous cognitive frameworks

Figure 1. Organizational Knowledge Environments: The Locus of Knowledge Diversity



Notes. The dashed lines represent production team boundaries and the solid lines represent firm boundaries. Distinct shapes represent inventors with different technological specializations. Two alternate firm-level approaches to organizing inventors on a team are depicted. Firm A represents a firm with a diffuse organizational knowledge environment consisting of high within-team knowledge diversity and low across-team knowledge diversity. Firm B represents a firm with a concentrated organizational knowledge environment with low within-team knowledge diversity and high across-team knowledge diversity.

and the lack of shared common building blocks among team members (Reagans and Zuckerman 2001, Srikanth and Puranam 2011). At the same time, the existence of knowledge diversity within any given team likely reduces the motivation (and perceived need) for seeking knowledge outside the focal team, even though doing so may be comparatively low cost from an accessibility standpoint (shared knowledge backgrounds across teams may facilitate knowledge exchange). The lack of shared common building blocks, together with the more siloed approach to knowledge recombination, likely lead to lower innovation quality under a diffuse organizational knowledge environment.

Hypothesis 1(a). *Firms characterized by a diffuse organizational knowledge environment (high within-team diversity, low across-team diversity) are associated with lower innovation quality.*

At the same time, this siloed approach to innovation may have a secondary effect that shapes the distribution of innovation outcomes across teams. Similar teams—each pursuing their own innovations while remaining relatively inert with respect to their surroundings (and other teams)—likely arrive at solutions with differential levels of effectiveness. Reframed from an evolutionary perspective, teams composed of similar genotypes are likely to exhibit different phenotypes when innovating in greater isolation, resulting in more variation in quality across teams. The diversity of combinations produced within the context of a firm-level selection environment can be conceptualized as being akin to an evolutionary mechanism of variation and mutation in speciation. This process likely results in more diversity in the fitness of teams and their resulting combination of ideas. However, only a smaller subset of these diverse, inert teams is likely to be fit given the (revealed) selection environment. As a consequence, under a diffuse structure, we expect greater inequality across teams with respect to innovation quality.

Hypothesis 1(b). *Firms characterized by a diffuse organizational knowledge environment (high within-team diversity, low across-team diversity) are associated with more unequal team contributions to their innovation output.*

The core operative mechanism underlying our prediction of lower innovation quality in Hypothesis 1(a) is the lack of a shared cognitive framework and work process within production teams as a result of greater within-team diversity. The ensuing coordination frictions stymie the efficacy of within-team diversity when inventors with different knowledge backgrounds engage with one another as part of the innovation process. These challenges to within-team coordination can be mitigated through a variety of organizational means (Aggarwal and Wu 2015). Shared team processes are one approach to the challenge of within-team

coordination in the presence of diverse cognitive frameworks. These processes arise when members of the same team informally learn by collaborating on prior inventions. A longer history of shared production experience by members of a production team allows for the emergence of routines and processes for coordination, thereby muting the negative effect of dissonant cognitive frameworks and work processes (Kotha et al. 2012). Consequently, this form of coordination through jointly shared prior experience may limit the detrimental innovation effects of inert teams under a diffuse organizational knowledge structure.³ We make the following hypothesis.

Hypothesis 2. *Factors that facilitate within-team coordination will positively moderate (i.e., reduce the negative effect of) Hypothesis 1(a).*

2.3. Concentrated Organizational Knowledge Environments and Firm-Level Innovation

Firms operating in a more concentrated organizational knowledge environment likely exhibit the opposite pattern to those in diffuse environments with respect to their innovation output. In a concentrated knowledge environment, the shared backgrounds of inventors provide for common individual starting points to the innovation process. This enables more immediate team productivity. At the same time, whereas the lack of common conceptual frameworks across production teams may limit access to more distant knowledge, the lack of knowledge diversity within the focal team implies greater benefits to spanning team boundaries (Rosenkopf and Nerkar 2001). Thus, there are likely to be high returns to sourcing knowledge from other teams within the firm. A concentrated knowledge environment therefore allows for a greater likelihood of across-team knowledge flows that in turn benefits innovation quality.

Hypothesis 3(a). *Firms characterized by a concentrated organizational knowledge environment (low within-team diversity, high across-team diversity) are associated with higher innovation quality.*

The higher level of motivation (and need) to engage externally to the team (but within the firm) likely increases flows of across-team knowledge and decreases the perceived threshold against which teams consult with others in the firm outside their own team. In other words, individuals may more readily consult with others outside their team when knowledge diversity is limited within the team, similar to the logic for when managers choose to consult with those outside their firm (Nickerson and Zenger 2004). Thus, despite knowledge being more concentrated within teams, team boundaries may effectively be more porous, enabling individuals to access areas of knowledge in which they themselves are not experts. Although

this more porous structure allows for higher mean levels of innovation quality as discussed earlier (due to the reduced need for coordination, paired with the benefits of accessing diverse knowledge as needed), it may also push firms toward greater overall homogeneity in innovation quality across teams. Given the motivation to engage externally, external consultation homogenizes differences and reduces expected quality variation across teams. Therefore, a more concentrated knowledge environment might engender more equal innovation outcomes among teams.

Hypothesis 3(b). *Firms characterized by a concentrated organizational knowledge environment (low within-team diversity, high across-team diversity) are associated with more equal team contributions to their innovation output.*

The core operative mechanism underlying our prediction of lower innovation quality in Hypothesis 2(a) is cross-team knowledge flows. These underlie the positive innovation quality outcomes expected from a concentrated organizational knowledge environment. In such a setting, there exists a shared (within-team) mutual understanding that allows for coordination, together with a rich organizational knowledge environment that requires boundary-spanning collaboration to reap knowledge-related benefits. When teams operate within a rich soup of (across-team) diversity, cross-fertilization of ideas allows a given team's inventors to benefit from the prior knowledge and experience of inventors on other teams. One implication of this is that firm characteristics that hinder the ease of across-team knowledge flows would reduce the positive effect of a concentrated team structure. For example, geographic dispersion—such as when the firm's inventors are spread across multiple locations—could increase the cost of across-team knowledge exchange (Haas and Hansen 2007, Lahiri 2010, Funk 2013) and thus reduce the positive effect of a concentrated knowledge environment. We make the following prediction.

Hypothesis 4. *Factors that increase the cost of knowledge accessibility across teams within a firm will negatively moderate (i.e., reduce the positive effect of) Hypothesis 3(a).*

Taken together, the arguments regarding differences in knowledge flows between diffuse and concentrated organizational environments suggest differences in intrafirm patterns of knowledge flows. Specifically, the relative proportion of knowledge on which a team builds that resides outside the team but within the firm would likely be higher in a concentrated organizational knowledge environment than in a diffuse organizational knowledge environment. Because the benefits of operating in a concentrated setting arise from diverse knowledge that resides outside

the team, teams in a concentrated environment more likely seek out and use the benefits of being situated in a soup of diverse knowledge. Doing so would imply building on a higher proportion of outside-the-team knowledge in their inventions than with inventors in a diffuse structure (which tend to be more inert and insular). This leads to our final prediction.

Hypothesis 5. *Concentrated organizational knowledge environments will draw on a greater proportion of within-firm external-to-the-team knowledge than will diffuse organizational knowledge environments.*

3. Methodology

3.1. Industry Setting

To empirically test our hypotheses regarding the firm-level innovation implications of heterogeneous intrafirm knowledge environments, we seek an industry in which innovation output is a key performance metric, where multiple invention teams within a firm is the norm, and where any given inventive team likely holds knowledge that can be of use to other teams, since teams operate within a shared overall knowledge space. The biotechnology industry satisfies these criteria and also has other beneficial characteristics: the industry is one in which knowledge-based resources are an important driver of innovation output, consistent with our objective of studying the link between the organization of knowledge-oriented human capital and firm-level innovation; and studying new venture evolution in this industry avoids issues of left-censoring, allowing us to track the same firms as they develop over time.

Patents are widely regarded as a key means of value appropriation in the biotechnology industry (Levin et al. 1987), and as such we can be more confident in relying on patent data to capture the relevant individual- and team-level characteristics that serve as precursors to the firm's innovative output. We rely on patent data to infer the structure of the firm's production teams for empirical observation. Biotechnology firms generally patent their inventions, and the resulting patent records—which contain names of individual inventors—reveal team composition to observers. For empirical purposes, production teams are defined as the set of inventors meriting attribution on a patent team. Patent data therefore has the primary benefit of enabling us to observe staffing in a large sample panel, which would otherwise be difficult to obtain systematically, especially across firms. In addition, the historical patent record allows us to construct detailed inventor career histories, allowing for comprehensive tracking of an inventor's technical experience both before and during their tenure at an in-sample firm. We use this inventive experience as a measure of the inventor's knowledge.

Although we leverage this data to observe a large sample of realized team structures, identifying teams in this way does impose some inference limitations. Our data reveal the ex post realized team structures associated with inventions, but not necessarily the ex ante team structures, as teams that did not successfully patent are censored (see Section 5 for more on this issue).

3.2. Data and Sample

To construct our sample, we seek a set of firms that is as homogeneous as possible, apart from the dimension of team organization, to make comparable and meaningful comparisons. Confining the sample to a single industry provides uniformity in interpreting firm-level objectives. Additionally, restricting the sample to venture capital-backed firms increases the commonality of the objectives and time horizons facing firms in the sample. Together, these factors reduce unobserved differences across firms, aside from the desired dimension of heterogeneity in team organization.

Our empirical sample is the universe of 476 venture capital-backed human biotechnology firms (SIC codes 2833–2836) founded between 1980 and 2000, as identified using the VentureXpert database. Our primary data set is an unbalanced firm-year panel in which firms are observed from their year of founding through either 2009 or, if sooner, their year of dissolution (a longer time window facilitates within-firm inferences). In addition to including all years in which the firm is privately held, we also include in our observation window firm-years following an initial public offering (IPO) or acquisition by another entity, together with controls for these ownership regimes.⁴ We use several sources to construct our variables. The Institute for Quantitative Social Science Patent Network database includes all U.S. Patent and Trademark Office data on patents applied for since 1975 (Li et al. 2014), allowing us to uniquely identify inventors associated with patents and to construct various measures of the production teams engaged in the creation of patents. Firm-year level attributes come from Deloitte Recap RDNA, Pharamaprojects, Inteleos, ThomsonOne, Zephyr, and SEC filings.

3.3. Dependent Variables

To measure the innovation performance implications of intrafirm organization (Hypotheses 1(a), 2, 3(a), and 4), our main dependent variable measures the number of forward citations received within a four-year postapplication window to the firm's patents in the focal firm-year. Forward citations are an accepted measure of innovation output (Trajtenberg 1990, Jaffe and Trajtenberg 2002), acting as a proxy for the economic value created by the patented innovations.⁵ A fixed citation window facilitates meaningful comparisons

across observation years. Without such a window, older patents would be upward biased in citation counts.⁶

To provide insight on the induced behavior underlying the aggregate firm-level innovation outcomes (Hypotheses 1(b), 3(b)), we also construct the dependent variable, *forward citations Herfindahl*, which measures the concentration of a firm's forward citations among its patents in a focal year. The measure is calculated using the standard Herfindahl index formula. Each patent generates a share of the forward citations (four-year postapplication window) to the firm's patents in a focal year. The share for a given patent is calculated as the forward citations to the given patent divided by the total forward citations to all of the firm's patents in the focal year. We then take the sum across the square of each of those shares to generate the firm-year level measure of *forward citations Herfindahl*.⁷ This measure ranges from 0 to 1, with a lower value indicating that forward citations come from a broad base of the firm's patents, and a higher value indicating that forward citations emanate from a more limited (concentrated) set of the firm's patents.

To test Hypothesis 5, we build a measure of the utilization of knowledge within the firm but outside of the team generating a given patent. *Firm nonteam self-citation ratio* is the count of backward citations by the firm's patents in the focal firm-year that do not cite patents by one of the citing patent team's inventors (but is a firm-level self-cite), divided by the total number of backward citations in the focal firm-year. This ratio normalizes the count relative to the overall volume of backward citations being generated by the firm, though the results we report later are also qualitatively the same for nonteam self-citation counts (levels). In other words, the numerator of this measure captures the volume of knowledge flowing inside the firm from one team to another, distinct team, and the denominator adjusts this measure for the total volume of knowledge the firm builds on in the firm-year.⁸

3.4. Main Independent Variables

Our three primary independent variables address the composition of technical knowledge diversity across and within teams. Before we describe these three independent variables, we first describe two intermediate indices, within-team knowledge diversity and across-team knowledge diversity, which are used to construct the three independent variables used in the regression analyses.

Our primary independent variables of theoretical interest are a combination of within- and across-team knowledge diversity, which we measure using individual inventors' career patenting experience. The intermediate *within-team knowledge diversity* measure captures the diversity of knowledge among different

inventors on a particular production team.⁹ To create this measure, we first measure the angular distance between the knowledge experience of each pair of inventors on a team (as described in further detail later as well as in Appendix Table A.1), and then take the mean of this value across all pairs of inventors on a team. The *within-team knowledge diversity* measure is then the average of this team-level value for all teams in a firm-year (our level of analysis). The *across-team knowledge diversity* measure captures the degree to which production teams within the firm differ from one another with respect to each team's aggregate knowledge. This variable is the average of the knowledge experience angular distance between all pairs of teams in the firm-year (as described in further detail later, as well as in the Appendix).

The angular distance between pairs of inventors (used to calculate *within-team knowledge diversity*) and pairs of production teams (used to calculate *across-team knowledge diversity*) are both based on the cosine similarity measure of Jaffe (1986). For all inventors at each year of their careers, we capture their knowledge experience with a class experience vector, consisting of the total experience the inventor has had patenting in each technology class. An inventor's total experience includes the inventor's entire history, and as such captures experience not only in the context of the current firm, but also in all prior firms. For a particular inventor, a given entry in the class experience vector thus represents the stock count of that inventor's patents in that particular technological class through the focal year, with the dimension of the vector being the total number of primary USPTO patent classes. We follow the cosine similarity measure of Jaffe (1986) by calculating the angular separation between the corresponding class experience vectors. Cosine diversity is defined as 1 minus the cosine similarity measure; the two diversity measures range from 0 to 1, where 1 is completely diverse (no overlap among class experience vectors) and 0 is completely homogeneous (full overlap among class experience vectors).

The *within-team knowledge diversity* measure is created by forming all possible dyads between inventors on a particular production team, calculating the cosine diversity for each dyad, averaging this for inventor dyads on a production team, and then aggregating to the firm-year level by taking the average for all patents in the firm-year. The *across-team knowledge diversity* measure is created by first summing the class experience vectors of the inventors on a given production team to create a single aggregate class experience vector for each production team. We then take the set of production teams of a firm in a year, form all possible dyads among these patents, calculate the cosine diversity measure among patents based on their respective aggregate class experience

vectors,¹⁰ and then calculate the average over all dyads of patents in the firm-year. We provide a detailed example of the calculation of the two knowledge diversity measures in the Appendix (Table A.1).

Taking these intermediate measures of across-team diversity and within-team diversity, we construct three main independent variables to represent the possible permutations of diversity using these two dimensions. We empirically categorize organizational knowledge environments into mutually exclusive categories to characterize and emphasize their distinction, and to map them to managerially relevant organization design decisions. We dichotomize these intermediate measures at the 75th percentile of their underlying distribution to capture the meaningful variation in the data, as shown in the scatterplot in Appendix Figure A.1 (for within-team diversity, the threshold is a value of 0.3, whereas for across-team diversity, the relevant value is 0.15) to empirically test the regimes of diversity discussed in our hypotheses.¹¹

We use *concentrated* (low within high across diversity) as an indicator (0/1) for firm-years where within-team knowledge diversity is below the 75th percentile of the sample and across-team diversity is above the 75th percentile. We define *diffuse* (high within low across diversity) (above 75th percentile and below 75th percentile, respectively) and *high within high across diversity* (above 75th percentile and above 75th percentile, respectively) similarly. All estimated coefficients on these three variables are relative to the excluded baseline case of low within low across diversity. We make comparisons relative to this baseline to understand how the locus of knowledge diversity, implied by the alternative organizational knowledge environments, shapes firm-level innovation outcomes. In addition, organizing production teams in a diffuse versus concentrated manner represents alternative ways of achieving organizational knowledge diversity (and as compared with the high within high across diversity environment, such potential managerial tradeoffs are not as salient, and so we do not devote much effort to interpreting that estimate).

3.5. Moderating Independent Variables

We construct a measure of lowered within-team coordination costs to test Hypothesis 2: prior within-team collaborative experience between different inventors on a team. For any two inventors i and $-i$, joint experience at time t is defined as the stock count of patents for which both inventors were on the same patent team at any point in their career through time t . The measure is constructed by averaging the joint experience among all dyads of inventors on the same patent team, which is then averaged over all patent teams in a firm-year.

To test Hypothesis 4, we build a measure for the accessibility of knowledge across teams within a firm.

We use *team geographic dispersion* to indicate the largest number of unique three-digit zip code prefixes among inventors on a production team, across the production teams patenting in the focal firm-year (as of this writing, there are 923 (over 40,000) three-digit (five-digit) zip codes in the United States). The basic U.S. Postal Service zip code format consists of five digits, where the first three digits represent a broader geographic region than the full five digits. Higher *team geographic dispersion* would make across-team consultation more difficult.

3.6. Control Variables

We employ a set of time-varying control variables (measured at the firm-year level) to account for residual differences beyond the time-invariant firm-level characteristics absorbed by including firm fixed effects across all models.

We include two control variables based on the patenting history of the firm and its inventors: *patent count* is the total number of patents applied for by the firm in the firm-year; and *total collaborative experience* is the average level of joint experience among all dyads formed from the full set of inventors patenting in a firm-year, constructed in a similar manner to within-team collaborative experience, described earlier. This latter control serves primarily as a counterpart to *within-team collaborative experience*, so that this independent variable of theoretical interest measures collaboration only within teams and does not errantly measure all collaborations within the firm.

We further account for time-varying characteristics of the firm that could correlate with both knowledge diversity characteristics and innovation. Collectively, these variables capture characteristics of firm quality and development stage that are particularly relevant in our industry setting of early-stage venture capital-backed biotechnology firms. The variables we create are: *age* of the firm in years since founding (from VentureXpert and public sources); *VC inflows stock*, which measures the cumulative venture capital (VC) investment into the firm (from VentureXpert); *strategic alliance stock*, which measures the cumulative stock of alliances in which the firm has been involved to date (from Deloitte Recap RDNA); and *active product*, which is an indicator for whether the firm has at least one active product in the U.S. Food and Drug Administration (FDA) pipeline (from PharmaProjects and Inteleos). We log *VC inflows stock* and *strategic alliance stock* in our regression models due to their skewed distributions. Although privately held is the baseline ownership regime, we also control for the firm's ownership using the post-IPO and post-M&A variables, hand-collected using archival news sources, as the ownership regime of the firm may influence both knowledge diversity characteristics and innovation

(Aggarwal and Hsu 2014). The variable *post-IPO* indicates that the firm has undergone an IPO in or before the focal year, and *post-M&A* is an indicator that the firm has been acquired in or before the focal year.

Finally, we control for top management team (TMT) characteristics that might influence both the firm's organization of knowledge diversity and its innovation output. Diverse top management teams are associated with greater innovation (Bantel and Jackson 1989), and TMT diversity may also correlate with the diversity of inventors that we study. For each firm, we manually collect data from public sources such as publicly viewable LinkedIn profiles and BoardEx to construct a full history of the TMT of each firm, focusing specifically on those holding C-suite titles (chief executive officer, chief technology officer, etc.). We construct intermediate Herfindahl concentration measures of various characteristics of the management team for each firm-year: age, job tenure, and educational duration diversity. Since these three measures are highly correlated, we average them into the composite measure *TMT diversity*.¹² Table 1 provides variable definitions and summary statistics, whereas Table 2 presents a correlation matrix of the variables.

3.7. Model Specification

We employ conditional fixed-effects Poisson models with robust standard errors in our main analyses at the firm-year level for our analysis of forward patent citations, and tobit models for our analysis of *forward citation Herfindahl* and *firm nonteam self-citation ratio*. The Poisson model accounts for the nonnegative count nature of the *forward citations* (Hausman et al. 1984, Hall and Ziedonis 2001).¹³ The tobit model accounts for the bounded nature of the associated dependent variables, which is asymptotically bounded at zero at the lower end and can take a maximum of one. To control for time-invariant firm characteristics, we include conditional firm fixed effects in the conditional Poisson model and traditional firm fixed effects in the tobit model (indicator variable for each firm). For both the panel Poisson and the tobit models, we include year fixed effects to control for year-over-year changes across firms. Together with the set of controls described earlier, these models facilitate the interpretation of our results as estimating within-firm, across-time effects. We report robust standard errors clustered at the firm level to account for possible within-firm error correlation across model years.¹⁴

4. Results

In Table 3, we present regression estimates of the models testing our main hypotheses. All models use a conditional fixed-effects Poisson estimation except for models (3-2) and (3-6), which use a tobit estimation with fixed effects. For the conditional Poisson models,

Table 1. Summary Statistics and Variable Definitions, Firm-Year Level of Analysis

Variable	Definition	Mean	SD
Dependent variables			
<i>Forward citations</i>	Total forward citations within a four-year window to patents filed in the focal firm-year	5.14	21.46
<i>Forward citation Herfindahl</i>	Herfindahl concentration of forward citations among patent teams in the focal firm-year	0.76	0.30
<i>Firm nonteam self-citation ratio</i>	Proportion of firm nonteam self-citations out of all backward citations in a focal firm-year	0.02	0.07
Main independent variables			
(1) <i>High within low across diversity</i> (Diffuse)	Indicator for heterogeneity within, homogeneity across, teams in technical experience	0.17	0.38
(2) <i>Low within high across diversity</i> (Concentrated)	Indicator for homogeneity within, heterogeneity across, teams in technical experience	0.13	0.34
(3) <i>High within high across diversity</i>	Indicator for heterogeneity within, heterogeneity across teams in technical experience	0.34	0.48
(4) <i>Within-team collaborative experience</i>	Average joint prior patenting experience between dyads of inventors on the same patent team	1.26	2.52
(5) <i>Team geographic dispersion</i>	Largest count of three-digit zip codes represented by the inventors on a patent team in a firm-year	1.05	1.67
Control variables			
(6) <i>Patent count</i>	Number of patents in a firm-year	2.24	7.93
(7) <i>Total collaborative experience</i>	Average joint patenting experience between all inventors	0.74	1.56
(8) <i>Age</i>	Years since firm founding	8.42	6.09
(9) <i>VC inflows stock</i>	Cumulative venture capital investment (\$M)	16.39	27.88
(10) <i>Alliance stock</i>	Stock count of strategic alliances	10.39	17.91
(11) <i>Active product</i>	Indicator for active product under FDA review	0.65	0.48
(12) <i>Post-IPO</i>	Indicator for IPO in firm history	0.32	0.47
(13) <i>Post-M&A</i>	Indicator for M&A in firm history	0.15	0.35
(14) <i>TMT diversity</i>	Average of Herfindahl TMT measures based on age, job tenure, and educational duration	0.08	0.19

the reported coefficients are incidence-rate ratios, which represent the exponentiated form of the regression coefficients. These coefficients can be interpreted as follows: for a unit increase in an independent variable, the incidence rate of the dependent variable would be expected to be scaled (multiplied) by the value of the estimated coefficient. Thus, a coefficient value less than one should be interpreted as a negative effect, whereas a

coefficient value greater than one should be interpreted as a positive effect. The tobit coefficients are reported in the standard manner (not as incidence-rate ratios).

All models include the three main independent variables: *high within low across diversity* (diffuse organizational knowledge environment), *low within high across* (concentrated organizational knowledge environment), and *high within high across diversity*.

Table 2. Pairwise Correlation Matrix of Independent Variables

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1.00													
2	−0.18	1.00												
3	−0.33	−0.28	1.00											
4	−0.13	0.23	−0.21	1.00										
5	−0.07	0.41	−0.33	0.58	1.00									
6	−0.09	0.31	−0.12	0.47	0.44	1.00								
7	−0.04	0.00	−0.22	0.77	0.42	0.08	1.00							
8	0.00	0.17	−0.35	0.06	0.08	0.08	0.02	1.00						
9	0.00	0.14	−0.23	0.20	0.21	0.09	0.14	0.11	1.00					
10	−0.07	0.30	−0.22	0.22	0.35	0.31	0.06	0.43	0.15	1.00				
11	−0.04	0.09	−0.06	−0.02	0.04	0.04	−0.04	0.34	−0.14	0.14	1.00			
12	−0.04	0.24	−0.31	0.21	0.32	0.19	0.12	0.39	0.21	0.45	0.27	1.00		
13	0.00	0.02	−0.13	−0.06	−0.11	−0.06	−0.04	0.30	0.06	0.07	−0.09	0.15	1.00	
14	−0.02	0.17	−0.16	0.19	0.27	0.16	0.11	0.11	0.30	0.29	−0.01	0.24	−0.09	1.00

Note. Variable numbering corresponds to Table 1.

Table 3. Regression Analysis

Model	(3-1)	(3-2)	(3-3)	(3-4)	(3-5)	(3-6)
Corresponding hypothesis	1(a)/3(a)	1(b)/3(b)	2	4	1-4	5
Dependent variable	<i>Forward citations</i>	<i>Forward citation Herfindahl</i>	<i>Forward citations</i>	<i>Forward citations</i>	<i>Forward citations</i>	<i>Firm nonteam self-citation ratio</i>
<i>High within low across diversity</i> (Diffuse)	0.624*** (0.106) [0.006]	0.225*** (0.049) [0.000]	0.490*** (0.080) [0.000]	0.713** (0.109) [0.027]	0.649*** (0.095) [0.003]	−0.040* (0.023) [0.087]
<i>Low within high across diversity</i> (Concentrated)	1.439*** (0.110) [0.000]	−0.155*** (0.036) [0.000]	1.467*** (0.111) [0.000]	2.096*** (0.345) [0.000]	2.026*** (0.322) [0.000]	0.033*** (0.013) [0.008]
<i>High within high across diversity</i>	0.458*** (0.069) [0.000]	−0.150*** (0.046) [0.001]	0.477*** (0.075) [0.000]	0.582*** (0.081) [0.000]	0.617*** (0.089) [0.001]	0.044** (0.017) [0.011]
<i>Within-team collaborative exp.</i>			1.041*** (0.013) [0.002]		1.042*** (0.012) [0.000]	
Diffuse × <i>Within-team collaborative exp.</i>			1.156*** (0.054) [0.002]		1.074* (0.047) [0.099]	
<i>Team geographic dispersion</i>				1.423*** (0.075) [0.000]	1.425*** (0.065) [0.000]	
Concentrated × <i>Team geographic dispersion</i>				0.863*** (0.041) [0.002]	0.879*** (0.040) [0.004]	
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Estimation model	Conditional Poisson	Tobit	Conditional Poisson	Conditional Poisson	Conditional Poisson	Tobit
Firms	300	305	300	300	300	328
Observations	4,843	3,926	4,843	4,843	4,843	2,380
Log pseudo-likelihood	−18,789.0	−1,527.8	−18,383.9	−16,050.8	−15,755.9	71.13

Notes. For conditional Poisson models, the reported exponentiated coefficients are incidence-rate ratios: a unit increase in an independent variable scales (multiplies) the dependent variable by the estimated coefficient. A coefficient value less (greater) than one represents a negative (positive) effect. For tobit models, reported coefficients are standard. Robust standard errors clustered at the firm level are shown in parentheses and *p*-values are shown in brackets.

p* < 0.10; *p* < 0.05; ****p* < 0.01.

The base category from which the other coefficients should be compared is low within low across diversity. All specifications include control variables, firm fixed effects, and year fixed effects.

4.1. Main Effects

In models (3-1) and (3-2), we find baseline results in support of Hypotheses 1(a), 1(b), 3(a), and 3(b). Model (3-1) considers the dependent variable of *forward citations*, and model (3-2) considers *forward citation Herfindahl*, where higher values indicate inequality in invention quality. A diffuse team knowledge structure (*high within low across diversity*) correlates with lower firm forward citations and higher inequality in invention quality, supporting Hypotheses 1(a) and 1(b), respectively. The diffuse team structure is associated

with 37.6% fewer forward citations to a firm's patents generated in a year. On the other hand, the concentrated team knowledge diversity structure (*low within high across diversity*) correlates with higher firm forward citations accompanied by more equal invention quality, supporting Hypotheses 3(a) and 3(b), respectively. The concentrated team structure is associated with 43.9% more forward citations to a firm's patents generated in a year. All of these effects are highly statistically significant (*p* < 0.01).

4.2. Mechanisms

In model (3-3), we add a moderating interaction term to test Hypothesis 2, a factor that facilitates within-team coordination. We find that the base term of *within-team collaborative experience* is positively associated

with *forward citations*. More importantly, we find that *within-team collaborative experience* positively moderates (makes less negative) the negative effect of a diffuse team structure ($p < 0.01$), supporting Hypothesis 3. One additional unit of past patenting experience between two team members is associated with 15.6% more *forward citations* for diffuse structures; expressed differently, a standard deviation increase in *within-team collaborative experience* is associated with 44.0% more forward citations for diffuse structures.¹⁵

In model (3-4), we find support for Hypothesis 4, the moderating effect of across-team knowledge accessibility on concentrated team structures. The base coefficient on *team geographic dispersion* is negative. When interacted with the concentrated team structure variable, we find a negative moderating effect (i.e., an effect that makes the main effect less positive) of *team geographic dispersion* ($p < 0.01$). An additional three-digit zip code of an inventor residence for the most dispersed patent team in a firm-year is associated with 13.7% fewer forward citations for concentrated structures; alternatively, a one standard deviation increase in *team geographic dispersion* is associated with 19.3% fewer forward citations. Model (3-5) includes both moderators in the same specification, with results consistent with (3-3) and (3-4), while preserving the main direct effect results of diffuse and concentrated organizational knowledge structures.

Finally, model (3-6) demonstrates support for Hypothesis 5 by showing that concentrated organizational knowledge environments draw on a greater proportion of within-firm external-to-the-team knowledge than diffuse organizational knowledge environments (with the opposite pattern for diffuse structures). We find that the concentrated structure is positively associated with *firm nonteam self-citation ratio* ($p < 0.01$), whereas the diffuse structure is negatively related ($p < 0.10$).¹⁶

4.3. Endogeneity Considerations

An endogeneity concern for our study is that there is an omitted variable that relates to firm quality that could correlate both with how knowledge diversity is organized within the firm (the diffuse and/or concentrated structure) and with innovation quality (the outcome variable). For example, a superior management team, in ways not captured by our *TMT diversity* control variable, may be able to impose a particular structure of knowledge diversity within the firm, while at the same time influencing the firm's innovation output quality in ways independent from our measures of knowledge diversity structure. More generally, the firm may select a structure due to an unobserved or unmeasurable factor that also drives innovation outcomes. We cannot fully exclude these possibilities, as we do not have an exogenous driver of organizational knowledge environments. However,

the inclusion of firm fixed effects mitigates the role of unobserved, time-invariant factors, and the multiple sets of controls further address factors that vary over time.

The second endogeneity concern is that of reverse causality between the dependent variable of innovation output and the main independent variables—that is, the possibility that firm innovation output drives the organizational knowledge environment structure. This reverse effect could be the case if firms with more promising innovation possibilities choose to select more or less diverse team structures to implement the innovation. Alternatively, firms with more innovative output may be able to become more diverse because they have greater resources from their innovation and can choose to be diverse, whereas other firms may be constrained from doing so. The antecedents of production team formation are an important research agenda in their own right, an issue we touch upon in our concluding discussion.

5. Discussion

In contrast with much of the extant literature on diversity and innovation, which focuses on the team- or subunit-level implications of knowledge diversity, we show that knowledge diversity arising from the firm-level organization of inventive human capital can play a key role in influencing the efficacy of a firm's knowledge-generation processes. We do so by introducing the concept of across-team knowledge diversity, which stands in contrast to the comparatively well-developed construct of within-team knowledge diversity. The across-team knowledge diversity concept opens new terrain in the ways in which organizations can design for knowledge diversity.

5.1. Theoretical Implications

The insights we develop have implications for several streams of the strategy literature, including work on intrafirm knowledge networks and the knowledge-based view of the firm. With regard to the former, a central theme of work on intrafirm networks is that the structure of social ties has implications for the degree to which individuals have access to and can use knowledge from distant parts of the firm (Tsai 2002, Nerkar and Paruchuri 2005, Grigoriou and Rothaermel 2013). Generally missing in this more structural account of intrafirm networks, however, are considerations regarding the characteristics of the nodes (e.g., the knowledge held by a particular inventor), as well as the ways in which these characteristics are distributed across the firm. Our results on across- and within-team knowledge diversity suggest that the distribution of inventive experience across invention teams is a key lever influencing firm-level innovation output due to the mechanisms of within-team coordination costs and across-team knowledge flows. Knowledge access and use stemming from the

distribution of human capital experience vis-à-vis production teams thus offers an explanation for heterogeneity in firm-level innovation that is distinct from network-related considerations. Future work might examine the degree to which these two sets of factors interact with one another to inform our understanding of firm-level innovation output.

Our work also has implications for the broader knowledge-based view of the firm (Kogut and Zander 1992, Grant 1996). A central theme of this literature is that competitive advantage, particularly in knowledge-based industries, is driven by differential firm-level abilities to organize internally held heterogeneous knowledge. Less attention has been paid, however, to the firm-level design principles necessary to achieve desired outcomes in this regard. As such, although the knowledge-based view has been deeply influential, it has made less headway with regard to the internal organizational correlates of innovation. One reason for this is that whereas the microfoundational processes underlying the knowledge-based view have been examined at the level of the individual team (substantial prior work has examined the knowledge-based correlates of team-level innovation [Ancona and Caldwell 1992, Hoegl et al. 2004]), limited progress has been made in developing a firm-level view that takes into account both within-team factors (e.g., coordination costs) together with across-team factors (e.g., cross-team knowledge flows). The insights from our study underscore the importance of considering firm-level team design in evaluating the efficacy of a firm's knowledge pool. Stated differently, when viewed at the level of the firm (versus the team), the organization design of a firm's inventive human capital can shape the efficacy of firm-level knowledge generation.

Finally, our paper has implications for recent work on the types of knowledge generated by organizations. Recent work by Pontikes and Barnett (2017) introduces the idea of "knowledge consistency," with the thesis that accretive knowledge (at the organizational level of analysis) is favored in the relevant product space. Two interlinked elements seem important to this concept: first, knowledge builds cumulatively and comprehensively; and second, the speed at which knowledge can be built matters. One extension of these ideas would be to examine the extent to which knowledge consistency can be achieved at the production team level, consistent with our focus on the locus of knowledge within firms. In particular, the organizational production team lens might be a way to study whether knowledge consistency operates mainly at the level of the firm or at the level of the production team, with implications for the knowledge accumulation process, particularly when the organizational environment shifts.

5.2. Limitations

In our conceptual development and empirical analysis, we abstract away from the antecedents of production team formation, instead focusing primarily on the innovation consequences of alternate patterns of knowledge diversity organization. Conceptually, production teams may arise from some combination of managerial fiat and self-organizing activities by the inventors themselves. In hierarchical organizations, managers may explicitly assign formal teams. Alternatively, production teams can self-organize: team members may organically search for team members, and then decide to work together. There is a limited literature on the process and outcome of self-organization, but such situations can be conceptualized either as informal working relationships that form "at the water cooler," or in a more intentional way, as in agile or scrum development teams, which are common in software industry settings. In a recent paper, Mortensen and Haas (2018) review the changing nature of team boundaries, arguing that teams have become more fluid, with increasingly greater degrees of overlap. Other recent papers examine the team performance consequences of such fluid configurations (Edmondson 2012, Valentine and Edmondson 2014).

We should be clear that our empirical results should not be interpreted as causal. Our main goal in this piece is to introduce and describe empirical patterns related to two ways of organizing knowledge diversity in firms (diffuse and concentrated organizational structures), and to provide evidence about the operative mechanisms as they relate to firm-level innovation. Although we discussed in the prior section some endogeneity and interpretational issues of the study more generally, given that there is no prior literature on across-team knowledge diversity (a central input to the organizational knowledge regime we discuss), it is unlikely that managers or inventors would have strong priors regarding how to organize to maximize the benefits of knowledge diversity.¹⁷ Future follow-on work might more fully address the team formation process, however, by conducting field experiments in inventive firms, where production teams are randomly assigned with respect to each inventor's patenting experience across technology classes. In addition, studies employing in-depth qualitative methods might examine the microprocesses through which individuals access and use knowledge, and in so doing contribute further to our understanding of how the organization of diversity shapes the efficacy and concentration of innovation output within a firm.

5.3. Conclusion

Prior literature focuses primarily on access to diverse knowledge as an input to the knowledge

recombination process. We suggest that when conceptualized at the firm-level, knowledge diversity can be characterized along two distinct dimensions: within-team and across-team. The latter captures the extent to which knowledge diversity of a firm's inventors is distributed across, rather than within, production teams, thereby taking into consideration the firm's internal organization. Our theory development and results suggest that the alternate forms of inventive human capital organization that arise by considering across-team diversity can be differentially effective with respect to firm-level innovation, as well as with respect to the concentration of such innovation across production teams in the firm.

To conclude, by introducing to the literature the novel dimension of across-team knowledge diversity, we advance our understanding of the link between

the intrafirm organization of inventive human capital and firm-level innovation.

Acknowledgments

For helpful comments, the authors thank Bill Barnett (Associate Editor), Martine Haas, Dan Levinthal (Editor-in-Chief), Henning Piezunka, Tim Simcoe, two anonymous reviewers, and participants at the Academy of Management Meetings, Brigham Young University-Utah Winter Strategy Conference, DRUID Conference, Strategic Management Society Conference, Strategy Science Conference, West Coast Research Symposium, and London Business School Trans-Atlantic Doctoral Conference. Sean Nicholson and Simon Wakeman generously shared biotechnology product and alliance data. Charles Du and Arthur Halim provided excellent research assistance. The authors dedicate this work in memory of Arthur Halim (1994–2016).

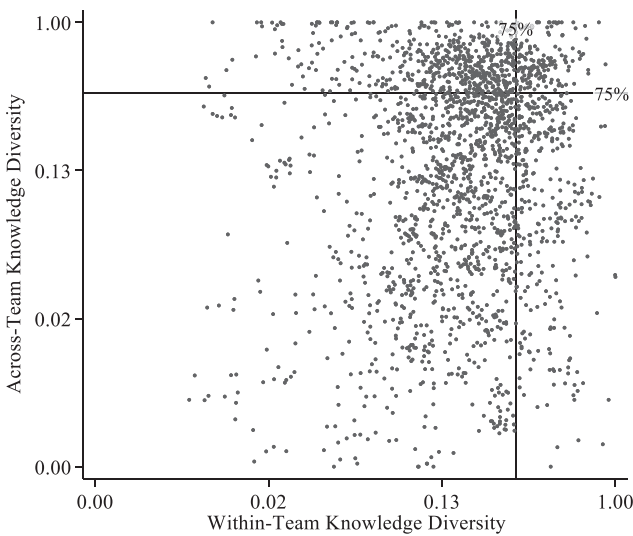
Appendix

Table A.1. Calculating the Across-Team and Within-Team Knowledge Diversity Measures

Class	Number of patents in each class						
	Team 1					Team 2	Team 3
	John	Paul	George	Ringo	Total	Total	Total
1	3	3	0	3	9	4	0
2	4	4	0	0	8	3	6
3	0	0	5	4	9	0	2

Within-team knowledge diversity of team 1				Across-team knowledge diversity of firm	
Pairings	Dyadic diversity $1 - \frac{A \cdot B}{\ A\ \cdot \ B\ }$		Pairings	Dyadic diversity $1 - \frac{A \cdot B}{\ A\ \cdot \ B\ }$	
John and Paul	0		Teams 1 and 2	$1 - \frac{12}{\sqrt{226}}$	
John and George	1		Teams 2 and 3	$1 - \frac{9}{5\sqrt{10}}$	
John and Ringo	16/25		Teams 3 and 1	$1 - \frac{33}{2\sqrt{565}}$	
Paul and George	1				
Paul and Ringo	16/25				
George and Ringo	1/5				
<i>Within-team knowledge diversity</i>	0.58		<i>Across-team knowledge diversity</i>	0.31	

Notes. We illustrate the calculation of across- and within-team knowledge diversity for a hypothetical firm. The top portion of the table shows the stock count of patents in each primary patent class for the members of team 1, and the total amounts for teams 2 and 3. The lower left portion of the table shows the calculation for the *within-team knowledge diversity* of team 1, where the dyadic diversity, which is 1 minus the cosine similarity of each dyad, is calculated for each dyad of inventors, and then averaged across all dyads. The lower right portion of the table shows the calculation of *across-team knowledge diversity* for the firm, where dyadic diversity is calculated for each dyad of teams and then averaged across all dyads.

Figure A.1. Scatterplot of Across-Team Diversity and Within-Team Diversity

Endnotes

¹ We focus our theorizing on input knowledge that is available within the firm (abstracting from input knowledge available from outside the firm). However, in our empirical analyses we control for the firm's use of this external knowledge by including covariates related to firm-level knowledge capabilities, which broadly capture the firm's absorptive capacity (Cohen and Levinthal 1990).

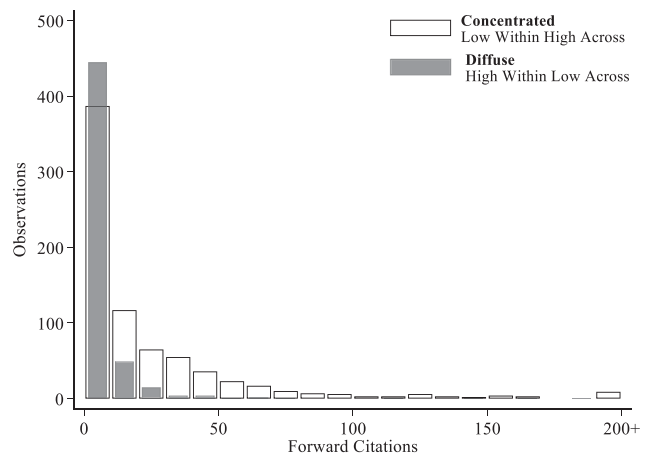
² A mix of factors comes into play when determining the set of individuals who directly contribute to a production team. These factors include both individual agency and managerial fiat, with the structures and processes leading to patterns of team formation arising as an outcome of the firm's organizational architecture (Joseph and Ocasio 2012). Although the determinants of this architecture may ultimately be subject to managerial control, numerous other factors (only a subset of which are observable) are likely to come into play. We thus abstract away from the determinants of the organization of inventive human capital, focusing instead on the implications of alternative such structures for firm-level knowledge production. We discuss this point further in the concluding section of the article.

³ Since these within-team coordination costs are not likely to substantively impact the selection environment, we do not investigate their relationship with the equality of innovation among teams.

⁴ To collect patent data on a firm's post-M&A years, we follow the procedure outlined in Aggarwal and Hsu (2014), which identifies a firm's inventors preacquisition and matches them to patenting activity by the same inventors in the acquiring firm postacquisition.

⁵ Our findings using forward citations within a four-year post-application window are robust to the inclusion or exclusion of self-citations to the inventions by the inventors on a patent team and self-citations to patents by the focal firm itself.

⁶ Given that some firms do not survive the entire period, there may be a concern that the exit of a firm has some effect on the pattern of forward citations to that exited firms' patents. To examine the empirical salience of possible survivorship bias, we conduct a separate analysis restricting the firm-year sample from founding up to four years before the end of a focal firm. For each firm in our sample, we manually identify survivorship status via press releases, news reports, and our previously mentioned databases. Since our forward citation-based dependent variables only rely on forward citations within a four-year window, this subsample ensures that we only consider firm-year observations where the firm survives through the

Figure A.2. Distribution of Forward Citations by Organization Knowledge Environment

four-year window for which we measure forward citations. The results are consistent with the main analysis presented in Table 3.

⁷ The formula for forward citations Herfindahl is

$$\sum_p \left[\frac{\text{forward citations}_p}{\sum_p \text{forward citations}_p} \right]^2$$

for all patents p by a firm in a year.

⁸ Sørensen and Stuart (2000) conceptualize organizational self-citation as a measure of building on own-knowledge, which naturally relates to the knowledge exploitation and exploration constructs. Rosenkopf and Nerkar (2001) consider self-citations in their study of organizational and technology domain-spanning firm behavior. Both papers distinguish between self-citing and non-self-citing patents as a way of measuring the extent to which firms build on their own prior knowledge as compared with exploring new domains. Our conceptualization starts from this lineage, but extends it for our purposes. We use self-citations to measure internal inventive activity, similar to these two papers. Building on Rosenkopf and Nerkar's (2001) notion of "internal boundary-spanning," we exploit the fact that self-citations list associated co-inventors. We use this as a window to observe the distinction between self-citation of inventions with and without at least one coinventor of a production team in common, a proxy for communication patterns and consultation within the firm. To our knowledge, this mechanism and measure is novel in the strategy literature.

⁹ We use a dyadic measure of diversity, which, as opposed to a measure based upon concentration (i.e., Herfindahl index) or variance, allows us to take the full range of an inventor's experience into account and to provide a complete measure of diversity among all the technological experience dimensions within a production team (i.e., a particular patent).

¹⁰ This measure contrasts with that of *within-team knowledge diversity*, where the cosine diversity measure is calculated among dyads of inventors within a team. In the case of *across-team knowledge diversity*, it is calculated among dyads of patents. We calculate *across-team knowledge diversity* across teams rather than across individuals (i.e., across the dyads of individuals in the firm) because we want a measure that captures the diversity of the firm in the context of the organization of teams within the firm. The inclusion of a firm-level diversity measure as calculated across all dyads of individual inventors in the firm does not change the sign or significance of the main independent variables, however.

¹¹ The results are also robust to a threshold corresponding to the 90th percentile, though high thresholds yield sparse data in the theoretically relevant independent variables (resulting in the majority of

observations falling into the omitted category (low across low within diversity).

¹² Our results are also robust to including any one of these measures individually.

¹³ A well-known issue with the Poisson model is the overdispersion one (the conditional mean in empirical data typically is smaller than the variance). To address concerns about overdispersion, we cluster (at the firm level) and implement robust standard errors in the panel conditional Poisson model. We also reestimate our models using a conditional fixed-effects negative binomial model instead, and we find consistent and statistically significant results to those reported.

¹⁴ We use the statistical software Stata/MP 15.1 to conduct all our empirical analysis. To estimate conditional fixed-effects Poisson regressions, we use the command *xtpoisson*, with the options of *fe* and *vce(robust)*. For *xtpoisson*, the *robust* standard error option by default clusters at the level of the panel variable, that is, firm, and is equivalent to explicitly specifying the variable to identify the panel (e.g., Wooldridge 2013). For tobit regressions, we use the *tobit* command with clustered standard errors by firm (*vce(cluster firm)*) with indicator variables for each firm.

¹⁵ For this and other interpretations of economic effect size in terms of standard deviation, we run an alternative model where we standardize the relevant variable and preserve the other characteristics of the presented model, that is, the same estimation method and other variables.

¹⁶ To rule out the concern that *firm nonteam self-citation ratio* has a mechanical relationship with concentrated or diffuse structures, that is, inventors in either environment have a different risk set of within-firm nonteam citations to make, we conduct an additional statistical test. The test evaluates whether these alternative organizational knowledge structures differentially relate to inventors' past co-patenting relationships in the form of exposure to peer patent class expertise through those relationships. We do not find a statistically significant difference in the co-patenting class exposure across concentrated versus diffuse structures, suggesting that our estimated relationships are not simply mechanical.

¹⁷ A different issue related to observed invention production teams is that we use patent data to study such teams. A common concern in studies using patent data to study organizations is censoring—that is, that some assembled teams did not successfully produce a patent, and therefore do not appear in our data. Under the assumption that the inventor teams that did not successfully receive granted patents would have received no forward citations as well if we were able to observe them, we explore the distribution of forward citations, segmented by diffuse and concentrated organizational knowledge structures, in the histogram presented in Appendix Figure A.2 (before any regression analysis). An empirical concern would be that the postregression positive relation of a concentrated organizational knowledge structure with innovation quality is upward biased due to the censored left density of forward citations by teams with a concentrated structure. Although we cannot know if censored knowledge structures are more or less likely to appear in the left tail of the forward citation distribution, from the data in our sample, we see that the overall concentrated structure effect is likely driven by the nonzero part of the distribution (whereas the opposite may be true for the diffuse structure, which has more density in the left tail of the distribution). Nevertheless, a caveat to our analysis is that censored team compositions may bias our estimates.

References

- Aggarwal VA, Hsu DH (2014) Entrepreneurial exits and innovation. *Management Sci.* 60(4):867–887.
- Aggarwal VA, Wu B (2015) Organizational constraints to adaptation: Intrafirm asymmetry in the locus of coordination. *Organ. Sci.* 26(1):218–238.
- Ancona DG, Caldwell DF (1992) Demography and design: Predictors of new product team performance. *Organ. Sci.* 3(3):321–341.
- Argyres NS, Silverman BS (2004) R&D, organization structure, and the development of corporate technological knowledge. *Strategic Management J.* 25(8–9):929–958.
- Bantel KA, Jackson SE (1989) Top management and innovations in banking: Does the composition of the top team make a difference? *Strategic Management J.* 10(S1):107–124.
- Carnabuci G, Operti E (2013) Where do firms' recombinant capabilities come from? Intraorganizational networks, knowledge, and firms' ability to innovate through technological recombination. *Strategic Management J.* 34(13):1591–1613.
- Cohen WM, Levinthal DA (1990) Absorptive capacity: A new perspective on learning and innovation. *Admin. Sci. Quart.* 35(1):128–152.
- Dixon NM (1994) *The Organizational Learning Cycle: How We Can Learn Collectively* (McGraw-Hill, London).
- Edmondson AC (2012) *Teaming: How Organizations Learn, Innovate, and Compete in the Knowledge Economy* (John Wiley & Sons, San Francisco, CA).
- Fleming L (2001) Recombinant uncertainty in technological search. *Management Sci.* 47(1):117–132.
- Fortunato S, Bergstrom CT, Börner K, Evans JA, Helbing D, Milojević S, Petersen AM, et al. (2018) Science of Science. *Science* 359(6379):eaao0185.
- Funk RJ (2013) Making the most of where you are: Geography, networks, and innovation in organizations. *Acad. Management J.* 57(1):193–222.
- Gerwin D, Moffat L (1997) Withdrawal of team autonomy during concurrent engineering. *Management Sci.* 43(9):1275–1287.
- Grant RM (1996) Toward a knowledge-based theory of the firm. *Strategic Management J.* 17(S2):109–122.
- Grigoriou K, Rothaermel FT (2013) Structural microfoundations of innovation: The role of relational stars. *J. Management* 40(2):586–615.
- Gruber M, Harhoff D, Hoisl K (2013) Knowledge recombination across technological boundaries: Scientists vs. engineers. *Management Sci.* 59(4):837–851.
- Haas MR, Hansen MT (2007) Different knowledge, different benefits: Toward a productivity perspective on knowledge sharing in organizations. *Strategic Management J.* 28(11):1133–1153.
- Hall BH, Ziedonis RH (2001) The patent paradox revisited: An empirical study of patenting in the U.S. semiconductor industry, 1979–1995. *RAND J. Econom.* 32(1):101–128.
- Hargadon A, Sutton RI (1997) Technology brokering and innovation in a product development firm. *Admin. Sci. Quart.* 42(4):716–749.
- Hausman JA, Hall BH, Griliches Z (1984) Econometric models for count data with an application to the patents–R&D relationship. *Econometrica* 52(4):909–938.
- Hoegl M, Weinkauff K, Gemuenden HG (2004) Interteam coordination, project commitment, and teamwork in multiteam R&D projects: A longitudinal study. *Organ. Sci.* 15(1):38–55.
- Hoisl K, Gruber M, Conti A (2016) R&D team diversity and performance in hypercompetitive environments. *Strategic Management J.* 38(7):1455–1477.
- Jaffe AB (1986) Technological opportunity and spillovers of R&D: Evidence from firms' patents, profits, and market value. *Amer. Econom. Rev.* 76(5):984–1001.
- Jaffe AB, Trajtenberg M (2002) *Patents, Citations, and Innovations: A Window on the Knowledge Economy* (MIT Press, Cambridge, MA).
- Jones BF (2009) The burden of knowledge and the “death of the renaissance man”: Is innovation getting harder? *Rev. Econom. Stud.* 76(1):283–317.
- Joseph J, Ocasio W (2012) Architecture, attention, and adaptation in the multibusiness firm: General Electric from 1951 to 2001. *Strategic Management J.* 33(6):633–660.

- Kogut B, Zander U (1992) Knowledge of the firm, combinative capabilities, and the replication of technology. *Organ. Sci.* 3(3): 383–397.
- Kotha R, George G, Srikanth K (2012) Bridging the mutual knowledge gap: Coordination and the commercialization of university science. *Acad. Management J.* 56(2):498–524.
- Lahiri N (2010) Geographic distribution of R&D activity: How does it affect innovation quality? *Acad. Management J.* 53(5):1194–1209.
- Leiponen A, Helfat CE (2011) Location, decentralization, and knowledge sources for innovation. *Organ. Sci.* 22(3):641–658.
- Levin RC, Klevorick AK, Nelson RR, Winter SG, Gilbert R, Griliches Z (1987) Appropriating the returns from industrial research and development. *Brookings Papers Econom. Activity* 1987(3): 783–831.
- Li GC, Lai R, D'Amour A, Doolin DM, Sun Y, Torvik VI, Amy ZY, Fleming L (2014) Disambiguation and co-authorship networks of the U.S. patent inventor database (1975–2010). *Res. Policy* 43(6):941–955.
- Mortensen M, Haas M (2018) Rethinking teams: from bounded membership to dynamic participation. *Organ. Sci.* 29(2):341–355.
- Nerkar A, Paruchuri S (2005) Evolution of R&D capabilities: The role of knowledge networks within a firm. *Management Sci.* 51(5):771–785.
- Nickerson JA, Zenger TR (2004) A knowledge-based theory of the firm—The problem-solving perspective. *Organ. Sci.* 15(6):617–632.
- Pontikes EG, Barnett WP (2017) The coevolution of organizational knowledge and market technology. *Strategy Sci.* 2(1):64–82.
- Reagans R, Zuckerman EW (2001) Networks, diversity, and productivity: The social capital of corporate R&D teams. *Organ. Sci.* 12(4):502–517.
- Rosenkopf L, Nerkar A (2001) Beyond local search: Boundary-spanning, exploration, and impact in the optical disk industry. *Strategic Management J.* 22(4):287–306.
- Sabbagh K (1996) *Twenty-First Century Jet: The Making and Marketing of the Boeing 777* (Scribner, New York).
- Schumpeter JA (1934) *Capitalism, Socialism, and Democracy* (Harper and Row, New York).
- Simon HA (1991) Bounded rationality and organizational learning. *Organ. Sci.* 2(1):125–134.
- Singh J, Fleming L (2010) Lone inventors as sources of breakthroughs: Myth or reality? *Management Sci.* 56(1):41–56.
- Sørensen JB, Stuart TE (2000) Aging, obsolescence, and organizational innovation. *Admin. Sci. Quart.* 45(1):81–112.
- Srikanth K, Puranam P (2011) Integrating distributed work: Comparing task design, communication, and tacit coordination mechanisms. *Strategic Management J.* 32(8):849–875.
- Trajtenberg M (1990) A penny for your quotes: Patent citations and the value of innovations. *RAND J. Econom.* 21(1):172–187.
- Tsai W (2002) Social structure of “coopetition” within a multiunit organization: Coordination, competition, and intraorganizational knowledge sharing. *Organ. Sci.* 13(2):179–190.
- Tushman ML (1977) Special boundary roles in the innovation process. *Admin. Sci. Quart.* 22(4):587–605.
- Valentine MA, Edmondson AC (2014) Team scaffolds: How meso-level structures enable role-based coordination in temporary groups. *Organ. Sci.* 26(2):405–422.
- Wooldridge JM (2013) *Introductory Econometrics: A Modern Approach*, 5th ed. (South-Western, Mason, OH).
- Wuchty S, Jones BF, Uzzi B (2007) The increasing dominance of teams in production of knowledge. *Science*. 316(5827):1036–1039.

Vikas A. Aggarwal is an associate professor of entrepreneurship and family enterprise at INSEAD, in Fontainebleau, France. His research uses empirical and computational modeling methods to study issues of strategy, organization design, and innovation in a new venture context as well as in situations of rapid industry change. His teaching focuses on private equity and venture capital. He received a PhD in management from the Wharton School of the University of Pennsylvania.

David H. Hsu is the Richard A. Sapp Professor of Management at the Wharton School of the University of Pennsylvania. His research and teaching are in the fields of entrepreneurship, innovation, and strategy. Hsu received his PhD in management from the Massachusetts Institute of Technology Sloan School of Management.

Andy Wu is an assistant professor of business administration in the Strategy Unit at Harvard Business School. His research and teaching are on strategy for entrepreneurial technology ventures. He received a PhD in applied economics from the Wharton School of the University of Pennsylvania, where he is a Senior Fellow at the Mack Institute for Innovation Management.