Fluid Teams and Knowledge Retrieval: Scaling Service Operations

Melissa A. Valentine
Stanford University
Stanford, CA 94305
Tel: 650.725.5676
mav@stanford.edu

Tom Fangyun Tan
Cox Business School, Southern Methodist University
Dallas, TX 75219
Tel: 214.768.1228
ttan@cox.smu.edu

Bradley R. Staats
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599-3490
Tel: 919.962.7343
bstaats@unc.edu

Amy C. Edmondson
Harvard Business School
Boston, MA 02163
Tel: 617.495.6732
aedmondson@hbs.edu

December 3, 2017

Acknowledgments

We thank Sambuddha Deb, Venkatesh Hulikal, Sowmya Narayanan, Ved Prakash, Alexis Samuel, and many other individuals at Wipro for their significant commitment of time and effort, which made this project possible. Martine Haas, Lamar Pierce, and Anita Tucker provided valuable comments on earlier drafts of this paper. We also thank Hau Lee, the associate editor, and the reviewers whose comments and suggestions greatly improved this work. This material is based on work supported by the National Science Foundation under Grant No. 0943210. Any errors are our own.

Abstract
To scale service operations requires retrieving knowledge across the organization. However, prior work highlights that individuals on the periphery of organizational knowledge networks may struggle to access useful knowledge at work. A knowledge repository has the potential to help peripheral individuals gain access to valuable knowledge because it is universally available and can be used without social interaction. However, for it to serve this equalizing function, those on the periphery of the organizational knowledge networks must actually use it, possibly overcoming barriers to doing so. In this paper, we develop a multi-level model of knowledge retrieval in teams to explore how individuals on the periphery of knowledge networks – due to inexperience, location, lack of social capital, gender, and role – access knowledge from a knowledge repository. Unexpectedly, we find that individuals whose experience and position already provide access to vital knowledge use a knowledge repository more frequently than individuals on the organizational periphery. We argue that this occurs because the knowledge repository – despite its appearance of equivalent accessibility – is actually more accessible to central than peripheral players. Thus, knowledge retrieval is not driven primarily by the need to overcome limited access to other knowledge sources. Rather knowledge retrieval is facilitated when actors know how to reap value from the knowledge repository, which ironically improves with increasing access to other sources of knowledge. We conclude that knowledge repository is unlikely to scale service operations without additional intervention.
Keywords: knowledge repository, knowledge management, scaling service operations, fluid teams

1. Introduction

In 2015, it was estimated that the Indian information technology (IT)-enabled services industry generated revenue of $146 billion, up from $105 million in 1989 (Athreye, 2005a; NASSCOM, 2015). A compound annual growth rate of over 30% for greater than 25 years is a remarkable industrial success story. The causes of success are myriad. Many important factors arose outside of the firms: changes in Indian governmental policy in 1991 encouraged IT exports; information systems became more prevalent in organizations; communication costs decreased due to innovations such as the internet; and globalization created new, global value chains (Arora, Arunachalam, Asundi, & Fernandes, 2001; Athreye, 2005b; Friedman, 2005).

However, the rise of the Indian IT-enabled services industry was not simply a story of being in the right place at the right time. The major players in the industry made significant investments in their own capabilities to gradually move their way up the global value chain (Arora & Asundi, 1999; Ethiraj, Kale, Krishnan, & Singh, 2005; Staats, Brunner, & Upton, 2011). When organizational performance depends on how knowledge is used, scaling a global service operation is particularly challenging. Their globally distributed employees often lack ready access to knowledgeable colleagues and struggle to retrieve knowledge across geographic and temporal distance (Cummings, Espinosa, & Pickering, 2009;

O'Leary & Cummings, 2007). In addition, many individuals in developing markets are relatively early in their careers, and are therefore not only removed from knowledgeable senior mentors, but also have less personal experience from which to draw upon in their daily work (Levenson, 2012; Ready, Hill, & Conger, 2008). Global organizations may want their employees to learn the most salient organizational information, but people early in their careers, or in remote offices, are likely to be on the periphery of knowledge networks (Singh, Hansen, & Podolny, 2010).

Information technology has been shown to offer promise in improving productivity (Bartel, Ichniowski, & Shaw, 2007; Brynjolfsson & Hitt, 1996). One such investment undertaken by each of the leading Indian IT-enabled services firms was the building of electronic knowledge repository (KR) systems. A KR offers a practical solution to the challenges of making knowledge available to people who might otherwise lack access to relevant expertise (Davenport & Prusak, 1998; Stein & Zwass, 1995). To create a KR, experts from across the firm codify best practices and useful insights, which are stored in an information system that is made available to all employees. These KRs are constantly available to every employee across time zones and continents and could thus function as a knowledge-access equalizer. A well-designed KR could help individuals access knowledge they would otherwise lack and help organizations manage the challenge of making knowledge available to employees independent of their social and geographic positions.

However, the presence of a KR alone will not solve the problem of knowledge access for those at the periphery of the organization unless it is used. Previous research has examined how the characteristics of the stored knowledge, such as employee perceptions of its quality, influence its use (Hansen & Haas, 2001; Kulkarni, Ravindran, & Freeze, 2006), but less is known about how knowledge network position, such as central vs. peripheral, affects knowledge retrieval. In this paper, we explore whether individuals on the organizational periphery take advantage of KRs, or KRs function more to enrich individuals whose experience and position already provide them better access to other knowledge sources.

We obtained archival data on knowledge retrieval at Wipro Technologies, a global outsourced provider of software services. Repository use was tracked for 10,703 individuals in 481 software project development teams on a per-click basis. We linked these empirical data with other Wipro databases including archival team and project characteristics. Our results show that – despite the seeming promise of a KR to integrate or equalize peripheral players – it instead enriches knowledge access for central players who are already well positioned. Our results suggest that knowledge retrieval is not simply an individual activity based on relative need, but is instead inhibited by an individual's network position. With these findings, we contribute to both the knowledge management and service operations literature. With respect to the former, we show that one's personal network and knowledge search within the defined confines of an information system may be encumbered by peripheral status (Singh et al., 2010). Using actual system use, we are able to show who is most limited. For the service operations literature, we contribute to the burgeoning topic of scaling operations. We find that good systems and processes are not enough by themselves, but instead, that system designers must address individual challenges if they are to scale successfully.

2. Setting

We developed understanding of our phenomenon at Wipro Technologies, a company operating in the software services industry. Wipro delivers software system development projects to its global client base. Team members rely on access to knowledge to complete their projects (Faraj & Sproull, 2000; Huckman & Staats, 2011; Huckman, Staats, & Upton, 2009), making this an ideal setting for understanding knowledge retrieval.

As mentioned previously, the Indian information technology (IT)-enabled services industry experienced dramatic growth during the 1990s and 2000s. As revenue grew, so too did headcount. From 1993 to 2014, headcount grew at a compound annual growth rate of over 18% (Heeks, 2015). At the same time, attrition of software engineers could run anywhere from 10% to 30% annually, depending on the company and the general environment. Companies, like Wipro, were faced with a very real challenge of how to scale their operations as they hired tens of thousands of employees per year. Wipro, similar to the other leading firms, invested significant money in identifying talented engineers – both at universities and from competing firms. Once they were brought into the organization, firms heavily invested in training these individuals and preparing them for the technical roles that they undertook.

However, Wipro and others still faced a considerable challenge when engineers were assigned to project teams in order to complete their work. Software development projects involve identifying a solution to client requirements, writing the software code to create the solution and then testing the final product (Boehm, 1981). To accomplish this work it is necessary to draw upon one's own experience and training. However, an individual may not have encountered a given type of problem previously, and so, needs to turn to others for help. Organizationally, it is sub-optimal to create new solutions for previously solved problems. Instead, the company would rather share known, viable solutions, so that others can save time and perhaps develop more innovative approaches.

Traditionally, individuals used their own ties to other employees or resources within a project to solve difficult problems or learn of alternative solutions. In an attempt to stay ahead (or at least even) with the strong competition from other software services firms (e.g., Accenture, Cognizant, IBM, TCS and Infosys), Wipro senior management sought to formalize their own knowledge management process. With a goal of delivering projects both efficiently and effectively, within the constraints of new employees and employee attrition, management focused on capturing and providing access to previously generated organizational knowledge through an electronic knowledge repository. The company had established a knowledge management initiative many years before, but at the time of this research, Wipro launched a new effort to enhance this initiative and invested substantial time and financial resources. Wipro enhanced the interface used for knowledge management (called KNet) and implemented analytic technology to enable the tracking of person-level use of the KR. All employees could download content from the KR and were encouraged to submit content. Submitted content was evaluated and solicited by a knowledge management team in order to maintain the intended quality standards of the system (the team functioned like the knowledge intermediaries described by Markus (2001)). Wipro did not dictate a specific policy on knowledge retrieval during the study period. The content of the KR included a limited amount of reusable software, but mainly consisted of documents detailing how the author of the document accomplished a specific task.

Describing how the KR could benefit work on projects, one project manager said, "There are many different types of documents on KNet. For example, there are case studies of prior projects that talk about benefits, problems, client value, new innovations, and best practices. There are also documents that explain how a specific aspect of a technology or domain works. These include details about solving particular problems, such as the flow of development, steps to follow, and examples." A different project manager commented, "In a project, we can go to KNet and use it to find best practices. We can look at case studies and see lessons learned and what issues different projects faced. All of this helps the team deliver better." As an example, a project manager explained, "On one project, we had a development activity that took two to three hours. We had to deploy code onto servers at a client location so it could be tested. I realized that by reordering the steps, I could create a simplified process that would do the same work in an hour and fifteen minutes. I posted a document explaining this, and then others [either on his project or other projects] could download it and save time."

Wipro employees learned about KNet during their initial training at the company. Moreover, the knowledge management (KM) team marketed the platform through emails to company personnel and through a link on the company's intranet. In the outreach efforts, the KM team focused on how the platform could help engineers on several business drivers, including shorter time to market, client responsiveness, and competency building. Although management believed that all employees could benefit from using the KNet system, they expected that the system would be particularly valuable to those employees that were not embedded in the firm's existing knowledge network, whom we label as peripheral players, since the system could provide an easy and straightforward way to access the knowledge of others.

We consider five different measures of peripheral status in the knowledge network based on prior literature and the global software services context: inexperience, remote location, limited familiarity with teammates, minority gender status, and front-line status. First, we consider the experience of team members. As noted earlier, the combination of company growth and turnover meant that a large number of engineers had limited firm experience. For example, approximately 67% of workers in our final sample had worked at Wipro for two years or less. Individuals were placed on projects as a function of how their experience set matched to the necessary work. It is important to note that software development projects within the software services space were not the same kind of development as might be found in writing a new software package or a new video game. Instead of being an entirely new creative process, it involved more problem solving to match a solution to the client's requirements. As a result, managers reported that the gap between individuals' needed knowledge and actual knowledge was either comparable for experienced and inexperienced workers, or slightly larger for inexperienced ones. In other words, experienced engineers might work on harder tasks, but they also had more knowledge; they were likely completing similar work to what they had done previously. Finally, managers noted that the technology for development projects was not changing so rapidly that experienced workers were growing obsolete. Rather, experienced workers were most prepared to work on these projects given their skill sets.

Second, we consider individuals' location. Wipro, like other software services, operated with a global delivery model where engineers were deployed around the world. However, the vast majority of engineers worked out of delivery centers in India (approximately 88% in our final sample). In India, engineers were typically located in large delivery centers with many other workers, creating numerous opportunities for knowledge sharing and interaction. Although the company had many facilities in India, at the time of our data collection, software development projects were typically staffed out of one location. Onsite engineers were deployed to the client locations in order to work on systems located at client sites. These individuals were completing similar work to what was being completed in India. Given that a majority of work was done in places like the US with 9½ to 13½ hour time differences, depending on the location and time of year, these onsite engineers were often isolated with little access to organizational knowledge networks. One project manager explained the need for a KR tool in such a setting by noting that, "When a team is spread out, then KR can be very beneficial. For example, the onshore team can go to KNet to check a code review checklist instead of having to contact someone on the offshore team. This makes code reviews faster and more accurate." During the time of our study, management reported that they tried to deliver as much work from India as possible, given the cost advantage. Engineers were sent onsite when systems required onsite access or when requested by clients. As a result, it is important to control for the client served when examining the effect of location on knowledge retrieval. Finally, we note that engineers located onsite are, on average, more experienced than engineers in offshore facilities. This occurs as brand new engineers are not typically sent onsite and instead, complete work in India before being eligible to go to a client site. Consequently, it is important to control for individual experience when examining the impact of location.

Third, we consider individuals' experience with other teammates. Given the general lack of firm-specific experience, it is not surprising that many individuals did not have experience with the same team members. In our final sample, the median engineer on a project had zero ties with other team members. This lack of interaction could serve as a barrier to knowledge search. However, although 51% of team members had not worked with anyone on a team previously, the other 49% had worked with at least one other team member; the latter may be better positioned to access knowledge.

Fourth, we examine minority gender status. Prior work notes that minority gender status may constrain knowledge search as individuals tend to seek knowledge from others who are of the same gender as them (Singh et al., 2010). In our final sample, only 26% of the engineers were female. Thus, we consider female as the minority gender status.

Finally, we consider role status and access to knowledge networks. Development projects typically had three roles – project engineer, middle manager, and project manager. While the latter two took on management responsibility, project engineers were responsible for execution. In our final sample, 84% of individuals were engineers (non-management). Individuals in management roles could be advantaged in the organizational knowledge network, as their status may help them secure access to knowledgeable experts and lead them to interact with these people. In the next section, we motivate the five hypotheses of our study.

3. Hypothesis Development

In this section, we investigate why the perspective that a knowledge repository might eliminate the knowledge gap between more and less connected individuals could be wrong. In particular, we theorize about how the conditions under which peripheral players' work both increases their need for knowledge access and also, ironically, and unfortunately, makes them less able to use a KR.

We begin by considering what it means to be core or peripheral within a firm's knowledge network. A long line of research has examined how networks, for knowledge or social relationships, often take a "small-world" structure whereby individuals can connect to others through a relatively small number of connections (Milgram, 1967; Travers & Milgram, 1969). Although subsequent research has supported the general small-world hypothesis, research has also revealed that unit-level differences may limit individuals' ability or inclination to cross over between different sub-groups (Fleming, King III, & Juda, 2007; Kleinfeld, 2002). Singh, Hansen and Podolny (2010) build upon this research by examining what individual factors may prevent an individual from having a small-world, and instead lead to long or ineffective searches for knowledge. Using data from a professional services firm, they theorize and find that individuals on the periphery of the organization's knowledge network – due to little experience, minority gender status, or poor connectedness – take significantly longer to reach an expert with a question than their more central counterparts do. The challenge, in part, is that these individuals not only do not know who knows what in the organization (Wegner, 1987), but also they pursue ineffective search strategies as they turn to others, like themselves, who also do not know where important knowledge resides.

A knowledge repository offers a potential solution to the lack of an organizational knowledge network. Since individuals may not know who has the requisite knowledge, or alternatively, are uncomfortable reaching outside of their own peripheral networks, they can instead use an electronic knowledge repository. This view highlights that *access* is the ability to acquire knowledge because of an open channel to it. Peripheral players have an open channel to the KR because it is universally available and does not require interpersonal interaction for use. Central players similarly have an open channel to the KR as a potential source of knowledge, but unlike peripheral players, they also have additional channels for potential knowledge flow because of ties to colleagues. Based on this conception of access, one could predict that peripheral players may be more likely than central players to use the KR: they may use the KR because it provides access where other access is unavailable, and central players may neglect the KR because of their abundant access to other knowledge sources.

However, a second perspective suggests a key challenge that might limit peripheral players. This perspective suggests that real access depends on understanding. That is, someone can access a calculus book by opening its pages, but the content might be inaccessible unless she understands math principles. As another analogy, someone might have access to a country with an open border, but the culture and accommodations might be inaccessible unless she understands the language. Both central and peripheral players enjoy the "open border" of a KR because of its networked interface and universal availability. However, central players have more than a terminal and a password— they are also embedded in locations and networks that also help them understand the language of the organization and the KR (see, for example Boh, 2008; Demian & Fruchter, 2006; Markus, 2001).

We propose that central players will be motivated to use the KR because they expect to be able to derive value from it due to their embeddedness in enabling social conditions (Bock, Sabherwal, & Qian, 2008; Fishbein & Ajzen, 1975). In contrast, peripheral players are likely left to find and understand relevant knowledge in the KR on their own. They may be intimidated by the metaphorical language barrier and less likely to get help accessing relevant knowledge. The KR provides access in the sense that peripheral players are able to make use of the KR whenever they want; however, this shallow access does not guarantee that the correct knowledge is identified or understood. For that kind of access, we suggest that social conditions that provide access to other sources of knowledge are needed.

We develop hypotheses for five conditions known to put people on the periphery of organizational knowledge networks – inexperience, remote location, limited familiarity with teammates, minority gender status, and front-line status – and argue that people working under each of these conditions will make less use of a KR than their more centrally positioned counterparts.

3.1 Individual Experience and Knowledge Retrieval

Despite this potentially real need to access knowledge, their limited experience may actually hinder knowledge retrieval for newer employees. First, they have less experience of their own on which to draw (Reagans, Argote, & Brooks, 2005). Their lack of personal experience may make much of the knowledge in the KR inaccessible. Knowledge retrieval is most effective when prompted by a well-developed question about how to do something specific (Markus, 2001). Inexperienced individuals may be struggling with questions like, "*What does this mean?*", or "*How does this relate to that?*", and not know how to pose a specific question in the technical language of the organization (Gray & Durcikova, 2005). Second, the less experienced workers tend to be on the periphery of organizational knowledge networks, and so have less access to knowledgeable colleagues (Singh et al., 2010). In other words, their position on the periphery of knowledge sharing networks means inexperienced individuals may not have help in searching for, or applying, what they find. Third, the colleagues to whom the inexperienced workers have access to, are likely to be inexperienced as well (Reagans, 2011; Singh et al., 2010), have equally little KR knowledge and be on the periphery of the knowledge network. Thus, we hypothesize:

HYPOTHESIS 1: *More experienced individuals will use the knowledge repository more than less experienced individuals.*

3.2 Location and Knowledge Retrieval

Individuals working in remote organizational locations also tend to be on the periphery of knowledge sharing networks. They interact with only a few colleagues face-to-face and confront significant challenges communicating and coordinating with their distant colleagues (Gibson & Gibbs, 2006; Hinds & Bailey, 2003; Martins, Gilson, & Maynard, 2004). Isolated team members, therefore, need rich modes of communication to support performance and collaboration with the rest of their team (Majchrzak, Malhotra, & John, 2005b; Malhotra, Majchrzak, Carman, & Lott, 2001). In other words, this disadvantaged locational position increases the need for access to non-local knowledge sources for individuals working in remote sites.

Despite acknowledging this potentially real need for knowledge retrieval, we argue that being located in the central rather than peripheral organizational offices is more conducive to knowledge retrieval for the following reasons. First, as with experienced individuals, individuals in the central offices have larger knowledge-sharing networks and thus, more social support that enables use. They can regularly interact with many colleagues because the central offices are usually much larger than remote offices, where only a few co-located colleagues may be available (O'Leary & Mortensen, 2010). Face-to-face interaction with a variety of colleagues enables easier discussions about new ways to solve or think about problems (Cramton, 2001), in addition to helping articulate specific questions, structure queries, and apply retrieved knowledge. Second, individuals in the central offices are supported in understanding the size and scope of the organization and how they relate to the knowledge stored in the KR. For example, central players typically work in a large, multi-space office. They pass by other project offices on their way to their own office space, thereby tacitly experiencing the scope of their organization (see for example Mortensen & Neeley, 2012). They will also have more non-project interactions in hallways and social spaces (Festinger, Schachter, & Back, 1963; Kiesler & Cummings, 2002). This may give them a uniquely comprehensive perspective on the expertise and the nature of the work in the firm (such as clients, industries, and technologies served), which may help them understand what is done and what is known in the organization (Orlikowski, 2002). In a central position, there is little need to create knowledge or solutions from scratch. Therefore, we hypothesize:

HYPOTHESIS 2: *Individuals located in the central organizational offices will use the knowledge repository more than individuals located in peripheral organizational offices.*

3.3 Team Familiarity and Knowledge Retrieval

A lack of social capital accumulated with team members can position individuals at the periphery of their team knowledge-sharing network and increase their need for knowledge access, but team members with more social capital are likely to be better supported in their knowledge retrieval. Individuals who are familiar with their teammates are more likely to trust and ask for help (Siemsen, Roth, Balasubramanian, & Anand, 2009). They will also have access to many different sources of knowledge within the team (Tucker, 2007), which makes it easier to use the KR, for example, to identify a closely related strategy that might not look related to someone with limited knowledge.

We also consider the influence of team-level familiarity on team members. This question considers how the overall level of social capital in the team influences the knowledge retrieval of a focal team member. We argue that overall team familiarity will support more knowledge retrieval by any focal team member. Team familiarity supports trust, information sharing, and effective coordination (Huckman & Staats, 2011; Pisano, Bohmer, & Edmondson, 2001). Familiar teams are more likely to have a shared understanding of task requirements (Balkundi & Harrison, 2006; Staats, 2012). Familiar teams are also more likely to know who knows what within the team and use this knowledge to coordinate their activities (Gino, Argote, Miron-Spektor, & Todorova, 2010; Lewis, Lange, & Gillis, 2005). Individuals on familiar teams will thus be better supported in knowledge retrieval. Familiar teams will help their team members to derive value from the KR, and will provide them with access to additional knowledge because of strong within-team networks, which in turn will support knowledge retrieval. Thus, we hypothesize:

HYPOTHESIS 3: *Individuals on teams with greater familiarity will use the knowledge repository more than those on teams with a lack of familiarity.*

3.4 Minority Gender Status and Knowledge Retrieval

The minority gender status of a worker within the team (in this case, female) may affect her using KR. First, being of the minority gender is likely to move the individual to the periphery of a company's knowledge network (e.g., Brass, 1985; Ibarra 1992). In particular, the minority gender is not only pushed to the less connected position in the workflow network (Podolny & Baron, 1997), but also often excluded from the social connections with the majority gender outside work, where people generally strengthen their bonds and communication (Kanter, 1993). For example, as noted by Singh et al. (2010), men may skip inviting women to happy hours after work because women tend to choose to directly go home for family reasons. Consequently, the minority gender may be unknowledgeable about how to use the KR or feel hesitant to consult other team members about the knowledge retrieval. Furthermore, being the minority gender may create homophily in terms of knowledge search strategy (Ibarra, 1997). Since the periphery position of the minority gender makes these workers less able to seek help with knowledge retrieval from the majority gender, they may turn to other minority gender workers, who equally have little knowledge about the KR. Thus, we hypothesize:

HYPOTHESIS 4: *Male employees will use the knowledge repository more than female employees.*

3.5 Front-Line Status and Knowledge Retrieval

The final dimension that we consider is that of the role within the team. Many teams, including the software development project teams that we analyzed, have a hierarchical structure whereby certain members (e.g. managers) are given rights and powers over the other team members. In general, given their greater status and knowledge, managers may be better positioned within a firm's knowledge network. For example, individuals must learn the responsibilities and tasks within their roles and so drawing on this fact, prior work finds that experience within each role may help improve team performance (Huckman et al., 2009). Given the progression from team member to manager, managers have served in the role as both managers and team members. Managers' advantageous knowledge network positions may help them better find knowledge as their role may enable their meeting clients or experts within the organization. Moreover, managers' increased status, due to their role, may make it easier for them to get a response from those whom they request knowledge.

Admittedly, team members are likely to have more use for a knowledge repository as it may help them close the knowledge gap that exists, as compared to a manager. However, as with the other dimensions, the better-positioned individual may actually use the repository more than the worse positioned individual. The greater awareness and understanding that a manager's role creates may also lead the manager to better understand what knowledge is available within a repository and to search more effectively to find it. Therefore, we hypothesize:

HYPOTHESIS 5: *Managers will use the knowledge repository more than individuals with front-line status (in this case, project engineers).*

4. Data

We received data from Wipro on all software development projects that occurred in a two-year period. The archival data captured individual knowledge retrieval and team project descriptors. Note that the knowledge retrieval data recorded how many unique downloads each individual completed on a given day during the two-year research period. Wipro did not document an identifier for the knowledge artifact viewed, so the specific content viewed is unknown. We matched the above data with demographic data on the 10,703 individuals that worked on software development team projects. We then dropped 130 projects each of which served only one client in the sample because we controlled for the fixed effects of the client being served in our empirical models. Controlling for the idiosyncratic and time-invariant aspects of a client (e.g., client involvement in the project) was important because they could affect knowledge retrieval. Our final dataset included information on the 481 software development projects that were started and completed during the study period, meaning we had comprehensive knowledge retrieval data for the entirety of each project's lifetime.

4.1 Dependent Variable

Table 1 provides summary statistics for the variables used in this paper.

Individual Knowledge Retrieval. The outcome variable, *Individual Knowledge Retrieval*, is calculated as $\log(\sum_1^n unique_n)$, where $unique_n$ equals the total number of unique knowledge artifacts accessed (i.e., downloaded) during a day and n denoted each day during the duration of the project. As a robustness check, we define a variable, *Anyuse*, which is equal to one if an individual ever uses the knowledge repository and is zero otherwise.

4.2 Independent Variables

Individual Experience. Consistent with prior literature examining human capital (Huckman & Pisano, 2006), we measure experience using an individual's firm-specific experience. We construct a variable, *Individual Experience* to capture the number of years an individual worked at Wipro prior to the start of the project. Hypothesis 1 predicts that more experienced individuals will be more likely to use the knowledge repository, so the coefficient on *Individual Experience* should be positive.

Individual Located in India. Wipro is an Indian organization, with its headquarters in India and the majority of its workers in India. Therefore, we define an individual who is located in an Indian office to be more central in the geographic network of the firm and accordingly create a variable, *Individual Located in India*, which is equal to one if an individual is located in an Indian facility, and is zero otherwise. Hypothesis 2 predicts that centrally located individuals will be more likely to use the knowledge repository, so the coefficient on *individual located in India* should be positive.

Team Familiarity. Project team members typically worked on one project at any given time and each team member was reassigned to new projects when the original project was completed. Team members also had different lengths of firm tenure. These dynamics created variability in the prior interactions among team members. To calculate *Team Familiarity*, we sum the number of projects that each unique dyad on the team completed together during the previous three years. We then divide the sum by the total unique dyads within the team to generate our variable (Reagans et al. 2005). Using a window of three years accounts for the potential decay of knowledge over time (e.g., Argote, Beckman, & Epple, 1990). The average project lasted for about seven months, so a three-year window matches the empirical context and allows us to include multiple cycles of projects.

An important question is whether or not to create a team-wide measure of the variable or to capture team familiarity between an individual and all other team members (i.e., the number of prior ties an individual has on the team). This individual-level measure is highly correlated with the team-level measure (correlation = 0.81), which may cause multicollinearity. In addition, prior work has relied upon the team-level measure (Huckman et al., 2009; Reagans et al., 2005). Hence, we choose to use the team-level measure to represent team familiarity. Hypothesis 3 predicts that familiarity with teammates will lead to an increased use of the knowledge repository, so the coefficient on *Team Familiarity* should be positive.

Gender. Since 26% of the individuals in our data are female, we define the minority gender in our setting as female. Accordingly, we construct an indicator, *Male*, which equals one if the individual is male and zero otherwise. Hypothesis 4 predicts that individuals of the majority gender will use the KR more than individuals of the minority gender, so the coefficient on *Male* should be positive.

Hierarchical Role. Individuals are classified into one of three roles within a project team. Project managers are responsible for leading the project. Depending on the project size, they may also do some technical work. Middle managers are leaders that assume a player/coach role. They lead sub-teams within the project, coach team members, and write code. Finally, project engineers are responsible for writing code and conducting the technical work of the project. We create an indicator, *Manager*, which equals one if the individual is a project manager or a middle manager and zero if she is an engineer. Hypothesis 5 predicts that managers will use the KR more than front-line individuals, so the coefficient on *Manager* should be positive.

4.3 Control Variables

We control for a number of other variables that may affect individual knowledge use.

Team Experience. We assess *Team Experience* by averaging the individual experience variable across all members on the team.

% of Team Located in India. As previously discussed, the software development teams that we study either work "offshore" at the company's facilities in India or "onsite" at client locations. We calculate the number of hours that were completed by the team at the Indian facilities and divide this value by the total number of hours worked by the team. Repeating the analyses with a variable calculated using the number of team members in each location, instead of the hours, we obtain the same pattern of results as those we report.

Male % on Team. To control for the gender dynamics on the team we calculate the number of hours completed on the team by males and then divide that value by the total number of hours worked by the team. Repeating the analyses with a variable calculated using the number of male team members, instead of the hours, produces similar results to those we report.

Manager % on Team. To control for team effects from hierarchical role differences we calculate the number of hours completed on the team by managers and then divide that value by the total number of hours worked by the team. Repeating the analyses with a variable calculated using the number of managers, instead of the hours, produces similar results to those we report.

Project Scale. We control for the scale of the project because more complex projects may have had greater knowledge needs and resulted in more use of a knowledge repository. We capture the project scale using the kilolines of new source code (KLOC) written (MacCormack, Verganti, & Iansiti, 2001). Prior work showed that software can exhibit scale effects (Banker & Kemerer, 1989), so we log-transform the variable in our models.

Estimated Effort and Duration. Projects that involved more effort in terms of hours or workdays might have been more difficult and may have required greater use of the knowledge repository. To control for both of these potential effects, we include the logarithm of the project's estimated total person-hours and the logarithm of the project's estimated total days.

Contract Type. Wipro used either a fixed-price contract structure or a time-and-materials contract structure for its development projects. In the former, the payment was agreed prior to the start of the project, while in the latter, Wipro received a pre-specified rate for the hours that they worked on the project. Given the role of incentives in individual and team performance, we control for contract type. We include an indicator variable in our models that is set to one if the project is fixed-price and is zero for time-and-materials.

Software Languages: Number and Type. Different software languages may have different knowledge demands leading to different patterns of knowledge retrieval. Similarly, projects with multiple software languages (53% of projects) may have greater knowledge demands and lead to more knowledge retrieval. We control for the former by including indicator variables for the different languages used (C, C++, Java, query, markup, BASIC). We adjust for the latter with an indicator variable equal to one if a project has more than one software language and zero otherwise.

Technologies. Projects that used multiple classes of technologies (e.g., client server, e-commerce) could lead to more knowledge retrieval. Thus, we create an indicator variable that equals one if a project has more than one technology (10% of projects) and zero otherwise (90% of projects).

4.4 Analysis Strategy

Since we were unable to conduct a field experiment to study our question of interest, we turn to archival data analysis. In such analysis, an important concern is whether selection effects might exist that then drive the subsequent results. For example, if familiar individuals were placed on teams that had a higher need for knowledge retrieval, then we might find the hypothesized effects, but for the wrong reason (team selection, rather than knowledge use challenges). Similar problems could exist if any of our other variables exhibited selection effects. In discussing this challenge with firm management, they did not see selection effects as a concern for two reasons. First, factors such as familiarity were not used to staff team members and project managers did not have the ability to select team members. Second, although different types of projects might have different knowledge needs, we are focusing on one type (i.e., development), which provides us with a "clean" setting to examine the individual-related effect on knowledge retrieval. We further control for important factors of the project, such as project complexity, estimated effort, software language used, and technology to mitigate this selection bias concern. Though, our design does not permit us to rule it out completely.

Our study contains data at three levels: the individual, the team project, and the client account. That is, individuals are nested in projects (teams), and projects are nested in client accounts. To be more precise, individuals are cross-classified because the same individuals can be sequentially staffed within different teams. Consequently, the individuals within the same team and the same teams working for the same client are likely to have similar knowledge retrieval needs. In addition, our hypothesized independent variables may have similar effects on the knowledge retrieval for the members within the same team, for the same client account. Ignoring these within cluster correlations at different levels in traditional estimation procedures, such as ordinary least squares (OLS), may cause wrong standard errors and even inconsistent estimates (Cameron & Trivedi, 2005).

In order to address this issue, we adopt a hierarchical linear modeling (HLM) approach in our analysis. The HLM technique allows correlations between the residuals in the same cluster (either teams or client accounts). In addition, it works for the unbalanced data structure in our setting in that each cluster has a different number of observations. For these reasons, HLM is often used in social sciences to analyze data that has a natural hierarchical structure (Raudenbush & Bryk, 2002). In particular, it was applied in recent studies on teams and knowledge management, which our study contributes to (Kang, 2012; Ko & Dennis, 2011).

We try to utilize the strengths of our unique data set on individual knowledge retrieval and sequentially construct multi-level models to ensure the consistency of our hypothesis testing related to individual use within project teams. Similar to other studies using HLM (DeHoratius & Raman, 2008), we first estimate a null model (Model 1), which includes only random intercepts at the client account level (Level 3) and project level (Level 2). We then introduce the hypothesized independent variables into the model (i.e., *Individual Experience, Individual Located in India, Team Familiarity, Male, Manager*). Finally, we add the control variables (defined in Section 4.3), which completes our full model (Model 2).

The null model of individual knowledge retrieval for observation *i*, which belongs to project *j* and client account *k* is defined as follows:

$$Individual\ Knowledge\ Retrieval_{ijk} = \alpha_0 + \beta_{00j} + \chi_{00k} + \varepsilon_{ijk}, \qquad (1)$$

where $\beta_{00j} \sim N\ (0, \sigma_j)$ represents the random main effect of project/team, and $\chi_{00k} \sim N\ (0, \sigma_k)$ is the random main effect of client account.

Our full model is specified as follows:

$$\begin{aligned} \textit{Individual Knowledge Retrieval}_{ijk} = {} & \alpha_0 + \beta_{00j} + \chi_{00k} + \lambda_1 \textit{Individual Experience}_{ijk} + \\ & \lambda_2 \textit{Individual Located in India}_{ijk} + \lambda_3 \textit{Team Familiarity}_{ijk} + \\ & \lambda_4 \textit{Male}_{ijk} + \lambda_5 \textit{Manager}_{ijk} + \delta_1 \textit{Team Experience}_j + \\ & \delta_2 \textit{\% of Team Located in India}_j + \delta_3 \textit{Male \% on Team}_j + \\ & \delta_4 \textit{Manager \% on Team}_j + \delta_5 \textit{Project Scale}_j + \delta_6 \textit{Estiamted Effort}_j + \\ & \delta_7 \textit{Estimated Duration}_j + \delta_8 \textit{Contract Type}_j + \delta_9 \textit{Software Language}_j + \\ & \delta_{10} \textit{Technologies}_j + \varepsilon_{ijk} \qquad (2) \end{aligned}$$

where $\lambda_1 - \lambda_5$ are random individual-level coefficients, varying across different projects (except for λ_3, which varies across client accounts because it is a team-level measure). We specify that these coefficients are random instead of being fixed because the likelihood ratio tests reject the null hypothesis that the slope coefficients are zero, and because the random slopes allow individual heterogeneity. In addition, we assume $\delta_1 - \delta_{10}$ to be fixed project-level coefficients of the control variables for computation reasons because too many random coefficients rapidly increase the number of parameters to estimate (Rabe-Hesketh & Skrondal, 2008).

All the models are estimated using full maximum likelihood technique (Stata command: xtmixed). In addition, we estimate crossed random effects to allow individuals to be nested within both project and client account levels because an individual may be sequentially staffed in different projects/teams.

5. Results

Table 2 provides a summary of our results. Column 1 presents the null model, where we find that the individual level (Level 1) contributes to 80% of the total variance in knowledge retrieval, the project level (Level 2) 16%, and the client account level (Level 3) 4%. The significant intraclass correlation (80%) at the individual level makes it interesting to examine the effect of individual's network position on knowledge retrieval, which we hypothesize. Column 2 shows the hierarchical model with just the independent variables and Column 3 adds the control variables to provide a more complete test of our hypotheses. We interpret the coefficients from Column 3 because the results of the independent variables are consistent across the models.

Hypothesis 1 predicts that individual experience would be positively associated with knowledge retrieval, and we find support for this hypothesis: the coefficient on individual experience is positive and statistically significant (0.0421). A one standard deviation increase in individual experience (2.36) is related to approximately 10% increase from the average amount of knowledge retrieval. At the mean level of knowledge retrieval (1.82) that would correspond to an individual using the KR approximately 0.2 more times per project, controlling for everything else.

In support of Hypothesis 2, which predicts that individuals located in India, Wipro's central location, would use the knowledge repository more, we find a statistically significant and positive coefficient of the individual's location in India (0.2279). An individual based in India would use the KR 22.79% more than an individual not based in India, which translates to about 0.4 times more than the current mean KR usage.

Turning to Hypothesis 3, team familiarity, we find that individuals on familiar teams are more likely to use the KR system than those on less familiar teams because the coefficient is significant and positive (0.2274). Interpreting this coefficient, we estimate that a one standard deviation increase in team familiarity (0.7) is related to a 15.9% increase in individual knowledge retrieval. At the mean level of knowledge retrieval on a project, that would correspond to an individual using the KR close to 0.3 times more per project.

Next, Hypothesis 4 suggests that members of the majority gender, males in this case, would be more likely to use the KR. The coefficient on *Male* turns out to be statistically indistinguishable from zero, so our hypothesis is unsupported.

Finally, supporting Hypothesis 5, which posits that managers will use KR more than front-line team members, we find that the coefficient of *Manager* is significant and positive (0.1614). In other words, on average, managers would use the KR 16.1% more frequently than front-line team members. This is close to 0.3 times at the mean level of knowledge retrieval per project.

Looking briefly at our other control variables, we see that the other team-level measures turn out to be statistically insignificantly related to knowledge retrieval except the following: *Male % on Team, Project Scale, Estimated Effort, Cplus,* and *Technologies.* For *Male % on Team,* although the result is partially statistically significant at 0.1 level, it is consistent with the idea that the minority gender (females) may use the KR more when surrounded by more of their own group members as this could provide them with the social support necessary to use the system (Gibson & Vermeulen, 2003). For *Project Scale, Estimated Effort, CPlus,* and *Technology,* the significant results provide evidence that supports that larger, demanding and complex projects create conditions for more knowledge retrieval, which is within our expectation.

6. Robustness Checks

A key question in an empirical paper is whether or not different modeling choices might have generated qualitatively different results. Therefore, we repeat our main model using a count outcome variable, a non-hierarchical model, and a dichotomous outcome variable in Table 3. Accordingly, Columns 1, 2, and 3 provide our robustness checks with the negative binomial, linear regression with client account fixed effects, and logistic regression models, respectively. Examining the coefficients, we find that each of our supported hypotheses (Hypothesis 1, 2, 3, and 5) continues to be supported. The coefficients in Columns 1, 2, and 3 are statistically and organizationally significant in each case and congruent with the results in Table 2.

An additional question arises with respect to our decision to capture familiarity as a team-level variable, rather than an individual-specific one. To address this concern, we repeat our main model and add a variable corresponding to *Individual Team Familiarity.* This variable is constructed by counting the number of prior projects between the focal individual and all others on the team and then dividing it by the number of team members minus one. Note that the individual team familiarity variable is highly correlated with our team familiarity variable (0.82). Whether we include the individual variable with the team one (shown in Table 3, Column 4) or only the individual variable (not shown), we find that *Individual Team Familiarity* turns out to show a statistically insignificant relationship with knowledge retrieval. Including it in the model does not meaningfully change the effect of team familiarity on knowledge retrieval, thus suggesting that our main results are robust.

Finally, for parsimony, we have used the mean values of the team variables to control for team-level effects. An alternative approach would be to include a dispersion measure, such as the Herfindahl index, which captures the dispersion of the underlying measure across the team. For example, a higher value of the Herfindahl index of team experience indicates that the experience within the team is concentrated within a smaller number of team members, while a lower value means that experience is more diffuse within the team. We introduce a Herfindahl measure for experience, team familiarity, location, gender and role, respectively. In Table 3, Column 5, we include the new measures to the original main model, and in Table 3, Column 6, we keep the Herfindahl measures and remove the team-level mean measures (except for team familiarity). In both cases, our key hypotheses continue to be supported.

7. Discussion and Conclusion

Scaling service operations is a challenge. In particular, developing the necessary skills within a workforce is difficult, and even more challenging for firms in developing markets. Hiring ready-trained individuals is often not an option. Instead, if the firm is to grow, it must create operational processes that support the development of the workforce. One such approach is to use a knowledge repository in order to address the gaps that may exist. However, we hypothesize that instead of the KR being used more by those individuals who are less central in the firm's knowledge network, just the opposite will be the case. Individuals who are more central in the knowledge network – due to experience, location, team familiarity, or role – have both awareness and the ability to access knowledge in a KR, and therefore, will use it more.

Our empirical results support our theory building. The key finding of this study is that use of a knowledge repository in a firm facing a scaling challenge is dominated by people who are operating at the center of organizational knowledge networks. We know from prior research that knowledge retrieval is also associated with objective team performance (Haas & Hansen, 2007). This presents a conundrum. Core players are expected to perform better to begin with because of their enabling work conditions (Cramton, 2001; Gardner, Gino, & Staats, 2012; Huckman et al., 2009). Our paper shows these conditions also support them in accessing additional knowledge resources. Thus, instead of the KR equalizing knowledge access across organizational members, the knowledge-embedded individuals became more embedded through using a KR. These results present crucial implications for two broad domains: team effectiveness and knowledge management.

Our study provides a critical perspective on team effectiveness for today's global firms, which increasingly deploy teams with fluid and distributed membership. For these teams to be effective, their leaders and managers must learn how to support coordination of work between unfamiliar and virtual teammates. KRs offer a potential plug-and-play technology solution by making knowledge available even to people with limited interpersonal knowledge access. Unfortunately, the KR does not overcome the taxing conditions faced by today's work teams to reach those at the organization's periphery. Inexperienced, remote, or front-line individuals, as well as those on unfamiliar teams, do not draw on the organizational resources provided in the KR. Taken alone, this technology solution fails to resolve the challenges facing globally distributed teams. Prior theory shows that these factors may be limits in the search of organizational knowledge (Singh et al., 2010) due to limited networks. The implication for knowledge management systems literature is that since it is an IT system, this should not be a problem (Alavi & Leidner, 2001). However, our results show that this is unequivocally not the case.

Our setting and analyses also provide a novel perspective on virtual teams. Much of the research on virtual teams has focused on supporting virtual team effectiveness (e.g., by establishing norms or developing trust) (Gibson & Cohen, 2003; Martins et al., 2004), but has not explored the relationship between virtual teams and their deploying organization. Even research on technology use in virtual teams has tended to focus on how information technology connects team members to *each other* (Hollingshead, 1996; Majchrzak, Malhotra, & John, 2005a), but less on how technology connects virtual team members to their organization as a whole. Workers spread across the globe are often working on similar projects for similar clients, and could thus, benefit greatly from each other's expertise. Therefore, keeping virtual workers connected to what the organization knows may provide strategic value. However, we show that dispersed individuals are less likely than their centrally located counterparts to make use of the organization's knowledge resources. Not only do they confront challenges accessing their teammates' knowledge, they also face challenges accessing the organization's electronic stores of knowledge. Future research could investigate what knowledge resources virtual teams draw on in their work and under what conditions they do so.

7.1 Limitations

Our findings have several important limitations. Although our data are archival and detailed, they did not allow us to see exactly which components were downloaded by individuals. This information could greatly enrich our understanding of the kinds of KR artifacts that are used, under what conditions, and how this influences performance. We argue that this limitation is not likely to bias our results in a systematic way (i.e., our hypotheses focus on amount of use rather than content of use), but it does prevent broader claims and understanding. In addition, our analysis is of one KR in one organization during a time when no active incentive programs were in place. There is considerable variation in how KM initiatives are structured (Boh, 2008; Davenport, De Long, & Beers, 1998); so, our results may not apply to all organizations with a KR. This setting allowed us to establish a baseline for the social conditions associated with knowledge retrieval. Moreover, Wipro is similar to many organizations in emerging markets that grow rapidly and are forced to scale their personnel in order to deliver a service. Therefore, we would expect to see similar results at other emerging market firms. In addition, the challenges that individuals on the periphery face are likely significant in firms based in developed markets; however, future research should explore this, in addition to different policies to manage KR.

An additional limitation comes from the archival nature of our study. Our empirical analysis relies on controlling for the factors in the data observable both to the researchers and the employees, and on qualitative discussions with firm personnel to indicate that unobservable factors are not biasing our results. Although this seems to be the case, we cannot guarantee it to be so. In other words, our study does not provide a precise mechanism for causal inference. For example, is it possible that experienced team members have seen their knowledge depreciate, such that they have greater knowledge demands and are therefore more likely to use the KR system to fill those needs, as compared to their inexperienced counterparts? As explained earlier, this is unlikely to be the case given the type of work that is conducted and the slow rate at which technology is changing. Nevertheless, future research should look to deploy causal methods, such as a field experiment to examine these questions in more detail.

7.2 Managerial Implications

Our findings offer both good news and a cautionary note for managers. AMR Research estimated that in 2007, the global market for knowledge repository and collaboration technologies was approximately $73 billion (Murphy & Hackbush, 2007). Prior research showed that *actual* use of a KR is in fact associated with objective team performance, which is optimistic news for those who have invested in these systems (Devaraj & Kohli, 2003; Haas & Hansen, 2007). However, our results also sound a cautionary note for managers. Despite the promise that KRs offer to global firms, in general, and developing market firms, in particular, we find that KRs enrich the knowledge access of central players in the organization; they are less effective at ensuring much needed access for those on the periphery. In other words, the KR provides a marginal performance gain for teams that would already be expected to perform well. Further work is needed to identify how managers can encourage productive knowledge retrieval – or offer other kinds of support – to individuals working in remote locations, who have limited organizational tenure, or who are in fluid teams where members do not know each other well. Just as research in productivity with information technology suggests that there is complementarity (e.g., Aral, Brynjolfsson, & Wu, 2012) – managers must make changes to fully benefit from the tools at hand – the same is likely the case for turning knowledge repositories into highly productive tools.

7.3 Conclusion

Scaling a service operation requires building and spreading knowledge throughout the firm. This challenge grows even more difficult when individuals are on the periphery of organizational knowledge sharing networks, due to inexperience, location, lack of social capital, or their front-line status. Although a knowledge repository is a potential tool to address such gaps, we show instead a pattern of use for individuals and teams working in more supportive conditions. Our results suggest that scaling service operations may depend critically on getting the conditions right, thus enabling people to support each other, to make use of available resources, and to perform more effectively.

8. References

Alavi, M., & Leidner, D. E. 2001. Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues. ***MIS Quarterly***, 25(1): 107-136.

Aral, S., Brynjolfsson, E., & Wu, L. 2012. Three-way complementarities: Performance pay, human resource analytics, and information technology. ***Management Science***, 58(5): 913 - 931.

Argote, L., Beckman, S. L., & Epple, D. 1990. The persistence and transfer of learning in industrial settings. ***Management Science***, 36(2): 140-154.

Arora, A., Arunachalam, V. S., Asundi, J., & Fernandes, R. 2001. The Indian software services industry. ***Research Policy***, 30(8): 1267-1287.

Arora, A., & Asundi, J. 1999. Quality certification and the economics of contract software development: A study of the Indian software industry, ***NBER Working Paper***.

Athreye, S. S. 2005a. The Indian software industry. In A. Arora, & A. Gambardella (Eds.), ***From Underdogs to Tigers: The Rise and Growth of the Software Industry in Brazil, China, India, Ireland, and Israel***: 7-40. London: Oxford University Press.

Athreye, S. S. 2005b. The Indian software industry and its evolving service capability. ***Industrial and Corporate Change***, 14(3): 393-418.

Balkundi, P., & Harrison, D. A. 2006. Ties, leaders, and time in teams: Strong inference about network structure's effects on team viability and performance. ***Academy of Management Journal***, 49(1): 49-68.

Banker, R. D., & Kemerer, C. F. 1989. Scale economies in new software development. ***IEEE Transactions on Software Engineering***, 15(10): 1199-1206.

Bartel, A., Ichniowski, C., & Shaw, K. 2007. How does information technology affect productivity? Plant-level comparisons of product innovation, process improvement, and worker skills. ***Quarterly Journal of Economics***, 122(4): 1721-1758.

Bock, G. W., Sabherwal, R., & Qian, Z. J. 2008. The effect of social context on the success of knowledge repository systems. ***IEEE Transactions on Engineering Management***, 55(4): 536-551.

Boehm, B. 1981. ***Software Engineering Economics***. New York: Prentice Hall.

Boh, W. F. 2008. Reuse of knowledge assets from repositories: A mixed methods study. ***Information & Management***, 45(6): 365-375.

Brynjolfsson, E., & Hitt, L. M. 1996. Paradox lost? Firm-level evidence on the returns to information systems spending. ***Management Science***, 42(4): 541-558.

Cameron, A. C., & Trivedi, P. K. 2005. ***Microeconometrics: methods and applications***: Cambridge university press.

Cramton, C. D. 2001. The mutual knowledge problem and its consequences for dispersed collaboration. ***Organization Science***, 12(3): 346-371.

Cummings, J. N., Espinosa, J. A., & Pickering, C. K. 2009. Crossing Spatial and Temporal Boundaries in Globally Distributed Projects: A Relational Model of Coordination Delay. ***Information Systems Research***, 20(3): 420-439.

Davenport, T. H., De Long, D. W., & Beers, M. C. 1998. Successful knowledge management projects. ***Sloan Management Review***, 39(2): 43-+.

Davenport, T. H., & Prusak, L. 1998. ***Working Knowledge: How Organizations Manage What They Know***. Boston: Harvard Business School Press.

DeHoratius, N., & Raman, A. 2008. Inventory record inaccuracy: An empirical analysis. ***Management Science***, 54(4): 627-641.

Demian, P., & Fruchter, R. 2006. An ethnographic study of design knowledge reuse in the architecture, engineering, and construction industry. ***Research in Engineering Design***, 16(4): 184-195.

Devaraj, S., & Kohli, R. 2003. Performance impacts of information technology: Is actual usage the missing link? ***Management Science***, 49(3): 273-289.

Ethiraj, S. K., Kale, P., Krishnan, M. S., & Singh, J. V. 2005. Where do capabilities come from and how do they matter? A study in the software services industry. ***Strategic Management Journal***, 26(1): 25-45.

Faraj, S., & Sproull, L. 2000. Coordinating expertise in software development teams. ***Management Science***, 46(12): 1554-1568.

Festinger, L., Schachter, S., & Back, K. W. 1963. ***Social Pressures in Informal Groups: A Study of Human Factors in Housing***. Stanford, CA: Stanford University Press.

Fishbein, M., & Ajzen, I. 1975. ***Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research***: Addison-Wesley.

Fleming, L., King III, C., & Juda, A. I. 2007. Small worlds and regional innovation. ***Organization Science***, 18(6): 938-954.

Friedman, T. L. 2005. ***The World is Flat: A Brief History of the Twenty-First Century*** (1st ed.). New York: Farrar, Straus and Giroux.

Gardner, H. K., Gino, F., & Staats, B. R. 2012. Dynamically integrating knowledge in teams: Transforming resources into performance. ***Academy of Management Journal***, 55(4).

Gibson, C., & Cohen, S. G. 2003. ***Virtual Teams That Work: Creating Conditions for Virtual Team Effectiveness***: Joey Bass.

Gibson, C., & Vermeulen, F. 2003. A healthy divide: Subgroups as a stimulus for team learning behavior. ***Administrative Science Quarterly***, 48(2): 202-239.

Gibson, C. B., & Gibbs, J. L. 2006. Unpacking the concept of virtuality: The effects of geographic dispersion, electronic dependence, dynamic structure, and national diversity on team innovation. ***Administrative Science Quarterly***, 51(3): 451-495.

Gino, F., Argote, L., Miron-Spektor, E., & Todorova, G. 2010. First, get your feet wet: The effects of learning from direct and indirect experience on team creativity. ***Organizational Behavior and Human Decision Processes***, 111(2): 102-115.

Gray, P. H., & Durcikova, A. 2005. The role of knowledge repositories in technical support environments: Speed versus learning in user performance. ***Journal of Management Information Systems***, 22(3): 159-190.

Haas, M. R., & Hansen, M. T. 2007. Different knowledge, different benefits: Toward a productivity perspective on knowledge sharing in organizations. ***Strategic Management Journal***, 28(11): 1133-1153.

Hansen, M. T., & Haas, M. R. 2001. Competing for attention in knowledge markets: Electronic document dissemination in a management consulting company. ***Administrative Science Quarterly***, 46(1): 1-28.

Heeks, R. 2015. Indian IT/Software Sector Statistics: 1980-2015 Time Series Data. https://ict4dblog.wordpress.com/2015/04/28/indian-itsoftware-sector-statistics-1980-2015-time-series-data/.

Hinds, P. J., & Bailey, D. E. 2003. Out of sight, out of sync: Understanding conflict in distributed teams. ***Organization Science***, 14(6): 615-632.

Hollingshead, A. B. 1996. Information suppression and status persistence in group decision making - The effects of communication media. ***Human Communication Research***, 23(2): 193-219.

Huckman, R. S., & Pisano, G. P. 2006. The firm specificity of individual performance: Evidence from cardiac surgery. ***Management Science***, 52(4): 473-488.

Huckman, R. S., & Staats, B. R. 2011. Fluid tasks and fluid teams: The impact of diversity in experience and team familiarity on team performance. ***Manufacturing & Service Operations Management***, 13(3): 310-328.

Huckman, R. S., Staats, B. R., & Upton, D. M. 2009. Team Familiarity, Role Experience, and Performance: Evidence from Indian Software Services. ***Management Science***, 55(1): 85-100.

Ibarra, H. 1997. Paving an alternative route: Gender differences in managerial networks. ***Social Psychology Quarterly***, 60(1): 91-102.

Kang, S., et al. . 2012. Research note—a multilevel analysis of the effect of group appropriation on collaborative technologies use and performance. ***Information Systems Research***, 23(1): 214-230.

Kanter, R. M. 1993. ***Men and women of the corporation: New edition***: Basic books.

Kiesler, S., & Cummings, J. (Eds.). 2002. ***What do we know about proximity and distance in work groups?*** Cambridge, MA: MIT Press.

Kleinfeld, J. S. 2002. The small world problem. ***Society***, 39(2): 61–66.

Ko, D.-G., & Dennis, A. R. 2011. Profiting from knowledge management: The impact of time and experience. ***Information Systems Research***, 22(1): 134-152.

Kulkarni, U. R., Ravindran, S., & Freeze, R. 2006. A knowledge management success model: Theoretical development and empirical validation. ***Journal of Management Information Systems***, 23(3): 309-347.

Levenson, A. 2012. Talent management: challenges of building cross-functional capability in high-performance work systems environments. ***Asia Pacific Journal of Human Resources***, 50(2): 187-204.

Lewis, K., Lange, D., & Gillis, L. 2005. Transactive Memory Systems, learning, and learning transfer. ***Organization Science***, 16(6): 581-598.

MacCormack, A., Verganti, R., & Iansiti, M. 2001. Developing products on "Internet time": The anatomy of a flexible development process. ***Management Science***, 47(1): 133-150.

Majchrzak, A., Malhotra, A., & John, R. 2005a. Perceived individual collaboration know-how development through information technology-enabled contextualization: Evidence from distributed teams. ***Information Systems Research***, 16(1): 9-27.

Majchrzak, A., Malhotra, A., & John, R. 2005b. Perceived individual collaboration know-how development through information technology–enabled contextualization. ***Information Systems Research***, 16(1): 9 - 27.

Malhotra, A., Majchrzak, A., Carman, R., & Lott, V. 2001. Radical innovation without collocation: A case study at Boeing-Rocketdyne. ***MIS Quarterly***, 25(2): 229-249.

Markus, M. L. 2001. Toward a Theory of Knowledge Reuse: Types of Knowledge Reuse Situations and Factors in Reuse Success. ***Journal of Management Information Systems***, 18(1): 57-93.

Martins, L. L., Gilson, L. L., & Maynard, M. T. 2004. Virtual teams: What do we know and where do we go from here? ***Journal of Management***, 30(6): 805-835.

Milgram, S. 1967. The small world problem. ***Phychology Today***, 1(1): 61-67.

Mortensen, M., & Neeley, T. 2012. Reflected Knowledge and Trust in Global Collaboration. ***Management Science***, In Press.

Murphy, J., & Hackbush, J. 2007. The knowledge management spending report, 2007–2008, ***AMR Research***.

NASSCOM. 2015. NASSCOM Industry Trends.

O'Leary, M. B., & Cummings, J. N. 2007. The spatial, temporal, and configurational characteristics of geographic dispersion in teams. ***MIS Quarterly***, 31(3): 433-452.

O'Leary, M. B., & Mortensen, M. 2010. Go (con)figure: Subgroups, imbalance, and isolates in geographically dispersed teams. ***Organization Science***, 21(1): 115-131.

Orlikowski, W. J. 2002. Knowing in practice: Enacting a collective capability in distributed organizing. ***Organization Science***, 13(3): 249-273.

Pisano, G. P., Bohmer, R. M. J., & Edmondson, A. C. 2001. Organizational differences in rates of learning: Evidence from the adoption of minimally invasive cardiac surgery. ***Management Science***, 47(6): 752-768.

Podolny, J. M., & Baron, J. N. 1997. Resources and relationships: Social networks and mobility in the workplace. ***American sociological review***: 673-693.

Rabe-Hesketh, S., & Skrondal, A. 2008. ***Multilevel and longitudinal modeling using Stata***: STATA press.

Raudenbush, S. W., & Bryk, A. S. 2002. ***Hierarchical linear models: Applications and data analysis methods***: Sage.

Ready, D. A., Hill, L. A., & Conger, J. A. 2008. Winning the Race for Talent in Emerging Markets. ***Harvard Business Review***, 86(11): 63-+.

Reagans, R. 2011. Close Encounters: Analyzing How Social Similarity and Propinquity Contribute to Strong Network Connections. ***Organization Science***, 22(4): 835-849.

Reagans, R., Argote, L., & Brooks, D. 2005. Individual experience and experience working together. ***Management Science***, 51(6): 869-881.

Siemsen, E., Roth, A. V., Balasubramanian, S., & Anand, G. 2009. The influence of psychological safety and confidence in knowledge on employee knowledge sharing. ***Manufacturing & Service Operations Management***, 11(3): 429-447.

Singh, J., Hansen, M. T., & Podolny, J. M. 2010. The World Is Not Small for Everyone: Inequity in Searching for Knowledge in Organizations. ***Management Science***, 56(9): 1415-1438.

Staats, B. R. 2012. Unpacking Team Familiarity: The Effects of Geographic Location and Hierarchical Role. ***Production and Operations Management***, 21(3): 619-635.

Staats, B. R., Brunner, D. J., & Upton, D. M. 2011. Lean principles, learning, and knowledge work: Evidence from a software services provider. ***Journal of Operations Management***, 29(5): 376-390.

Stein, E. W., & Zwass, V. 1995. Actualizing organizational memory with information systems. ***Information Systems Research***, 6(2): 85-117.

Travers, J., & Milgram, S. 1969. An experimental study of the small world problem. ***Sociometry***, 32(4): 425-443.

Tucker, A. L. 2007. An empirical study of system improvement by frontline employees in hospital units. ***Manufacturing & Service Operations Management***, 9(4): 492–505.

Wegner, D. M. 1987. Transactive memory: A contemporary analysis of the group mind. In G. Mullen, & G. Goethals (Eds.), ***Theories of Group Behavior***: 185-208. New York: Springer-Verlag.

Tables

Table 1. Summary statistics and correlation table of variables in the individual knowledge retrieval models (*n*= 13,470).

Variable	Mean	σ	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Individual Knowledge Retrieval	0.47	0.88													
2. Individual Experience	1.82	2.36	**0.12**												
3. Individual Located in India	0.88	0.33	**0.06**	**-0.18**											
4. Team Familiarity	0.29	0.70	**0.05**	**0.11**	**-0.03**										
5. Male	0.70	0.46	**0.03**	**0.11**	**-0.12**	0.00									
6. Manager	0.16	0.37	**0.10**	**0.37**	**-0.13**	**0.02**	**0.13**								
7. Team Experience	2.36	1.12	0.00	**0.14**	**-0.05**	**0.13**	0.01	**0.05**							
8. % of Team Located in India	0.86	0.14	**0.02**	**-0.04**	**0.28**	**-0.05**	0.00	**-0.04**	**-0.05**						
9. Male % on Team	0.75	0.13	-0.02	0.00	**-0.03**	**-0.06**	**0.15**	**0.04**	**0.12**	**-0.05**					
10. Manager % on Team	0.16	0.10	**0.05**	**0.10**	**-0.05**	**0.07**	**0.04**	**0.27**	**0.20**	**-0.15**	**0.15**				
11. Project Scale	3.32	1.98	**0.06**	**-0.03**	**0.01**	**-0.08**	-0.01	0.01	**-0.14**	-0.01	**-0.02**	**0.02**			
12. Estimated Effort	9.38	1.17	**0.11**	**-0.05**	**-0.03**	**-0.17**	0.00	0.00	**-0.14**	**-0.14**	**0.04**	**-0.02**	**0.32**		
13. Estimated Duration	5.63	0.62	**0.09**	**-0.04**	0.00	-0.18	0.00	**0.02**	**-0.20**	**-0.04**	**0.10**	**0.07**	**0.26**	**0.65**	
14. Contract Type	0.41	0.49	-0.01	**0.04**	-0.01	**-0.07**	**0.02**	0.00	**0.13**	**0.04**	**0.10**	0.02	**-0.02**	**0.03**	**-0.02**

Note. Bold denotes significance of less than 5%.

Table 2. Summary results of the regression of individual knowledge retrieval (n = 13,470).

	Dependent Variable: Individual Knowledge Retrieval		
	(1)	(2)	(3)
Individual Experience		0.0407***	0.0421***
		(0.0048)	(0.0048)
Individual Located in India		0.2216***	0.2279***
		(0.0258)	(0.0261)
Team Familiarity		0.2024**	0.2274***
		(0.0667)	(0.0684)
Male		0.0087	0.0146
		(0.0174)	(0.0177)
Manager		0.1599***	0.1614***
		(0.0274)	(0.0275)
Team Experience			-0.0136
			(0.0152)
% of Team Located in India			-0.0268
			(0.1210)
Male % on Team			-0.2017+
			(0.1152)
Manager % on Team			0.1051
			(0.1601)
Project Scale			0.0174+
			(0.0094)
Estimated Effort			0.0537*
			(0.0216)
Estimated Duration			0.0444
			(0.0365)
Contract Type			0.0012
			(0.0369)
Software Languages			-0.0289
			(0.0388)
C			0.0392
			(0.0441)
Cplus			0.0930*
			(0.0469)
Query			0.0056
			(0.0513)
Basic			0.0534
			(0.0664)
Technologies			-0.1312*
			(0.0614)
Constant	0.4949***	0.1462***	-0.4789*
	(0.0230)	(0.0347)	(0.2403)
Observations	13,470	13,470	13,470
Prob>Chi-sq	<0.001	<0.001	<0.001

Notes: +, *, ** and *** denote significance at the 10%, 5%, 1%, and 0.1% levels, respectively. Standard errors are shown in the parentheses.

Table 3. Summary results of robustness tests.

	Model: Negative Binomial	**Model: Fixed Effects Regression**	**Model: Logistic Regression**	**Personal Team Familiarity**	**Herfindahl within Teams**	**Herfindahls with Teams Excluding Average Effects**
	(1)	(2)	(3)	(4)	(5)	(6)
Individual Experience	0.0788***	0.0399***	0.1215***	0.0406***	0.0422***	0.0421***
	(0.0054)	(0.0034)	(0.0091)	(0.0035)	(0.0047)	(0.0047)
Individual Located in India	0.5096***	0.2252***	0.6711***	0.2230***	0.2244***	0.2195***
	(0.0521)	(0.0238)	(0.0682)	(0.0244)	(0.0255)	(0.0251)
Team Familiarity	0.0871***	0.0406**	0.1476***	0.0753***	0.2614***	0.2674***
	(0.0198)	(0.0149)	(0.0383)	(0.0187)	(0.0655)	(0.0663)
Male	-0.0376	0.0043	-0.0592	-0.0042	0.0115	0.0094
	(0.0334)	(0.0167)	(0.0453)	(0.0173)	(0.0177)	(0.0174)
Manager	0.3141***	0.1656***	0.4039***	0.1646***	0.1607***	0.1604***
	(0.0395)	(0.0219)	(0.0580)	(0.0228)	(0.0275)	(0.0275)
Personal Team Familiarity				0.0036		
				(0.0169)		
Herfidhal - Experience					0.8296***	0.7621***
					(0.1645)	(0.1622)
Herfindahl - Location					-0.5600***	-0.5700***
					(0.0730)	(0.0720)
Herfindahl - Team Familiarity					0.2714*	0.2880*
					(0.1288)	(0.1287)
Herfindahl - Gender					0.2285	0.0291
					(0.1769)	(0.1484)
Herfindahl - Project Role					-0.3044	-0.0115
					(0.4097)	(0.1144)
Other Project-level Controls	Included	Included	Included	Included	Included	Excluded
Constant	-3.9152***	-0.6405***		-1.0015***	-0.1365	-0.6323*
	(0.2738)	(0.1705)		(0.1079)	(0.4760)	(0.2466)
Observations	13413	13470	13409	13470	13470	13470

Notes: *, ** and *** denote significance at the 5%, 1%, and 0.1% levels, respectively. Standard errors are shown in the parentheses.