Learning to Cross Boundaries in Online Knowledge Communities: Fading of Surface-level and Rise of Deep-level Similarity with Experience

Elina H. Hwang · Param Vir Singh · Linda Argote

{elinah, psidhu, argote} @cmu.edu

Tepper School of Business & Center for the Future of Work, Heinz College

Carnegie Mellon University

5000 Forbes Avenue

Pittsburgh, Pennsylvania 15213

This research was supported by grant from the Future of Work Center at Carnegie Mellon University.

Learning to Cross Boundaries in Online Knowledge Communities: Fading of Surface-level and Rise of Deep-level Similarity with Experience

ABSTRACT

Many organizations have launched online knowledge forums to promote knowledge flow across boundaries. This paper theorizes and empirically tests whether employees transfer knowledge within or across boundaries and how the tendencies change as a function of a knowledge provider's experience in an online forum. We suggest that participants prefer to transfer knowledge to others with whom they share common ground and use joint characteristics with their communicating partners to assess the level of common ground. In particular, we propose that both surface-level (location and status) and deep-level (expertise) similarities of a dyad drive knowledge transfer. Further we propose that a knowledge provider's cumulative knowledge-sharing experience in an online forum moderates the relative effects of surface-level and deep-level similarities on knowledge transfer. Using panel data at an online knowledge forum of a large IT consulting firm, we find that similarity breeds connection in online forums. Additionally, we find that the effect of deep-level similarity on knowledge transfer increases whereas the effect of surface-level similarity decreases with knowledge providers' experience. That is, as participants gain experience, they provide knowledge less frequently to others at the same location and in the same status and more frequently to others with similar expertise.

Keywords: common ground; experience-based learning; knowledge transfer; online communities; surface- and deep-level similarity

INTRODUCTION

The ability to utilize existing knowledge is critical for an organization's success (Argote 1999, Grant 1996, Zander and Kogut 1995). Knowledge, however, is often insulated by boundaries that makes it challenging, if not impossible, to locate, acquire, and adopt knowledge across them. Many firms of today are organized on a global basis in order to take advantage of resources and markets around the world. In these geographically dispersed organizations, considerable knowledge is developed and accumulated locally, resulting in knowledge silos demarcated by a physical boundary. Further, employees tend to group by various demographic (e.g., age, gender) or social characteristics (e.g., hierarchical status) due to homophily effects (McPherson et al. 2001), which create additional boundaries along those attributes.

The web enters knowledge management as a solution for the bounded knowledge sharing. According to the survey by Forrester Research, 106 out of 119 CIOs in companies with more than 500 employees have deployed at least one type of social media[1] to aid their knowledge management (Framington 2007). The most popular social media adopted so far is web knowledge forums for employees (online forums hereafter). An online forum is an organizational version of *Yahoo! Answers*, a knowledge-networking online community for employees. Through online forums, employees can acquire solutions unavailable from their nearby colleagues. Just as *Yahoo! Answers*, most online forums provide interactive but asynchronous communication and rely on voluntary participation of employees in generating contents.

Mobilizing knowledge across boundaries has been touted as an advantage of online forums. First, because the Internet can bridge space and time, it makes it easy for geographically scattered workers to exchange ideas and solutions. Hence, an online forum is expected to break physical knowledge silos.

[1] A group of Internet-based applications that build on web 2.0 technologies, which allows the creation and exchange of user-generated content (Kietzmann et al. 2011). Internet forums, blogs, and social networking applications are examples of social media.

Second, knowledge sharers in online forums are not co-present and can communicate only verbally, so less or no social information (e.g., appearance) is available. The reduced level of social information shrinks the social distance among dissimilar people, resulting in increased interaction among them (Kiesler et al. 1984, Sproull and Kiesler 1986, 1991).

Thrilled by the great potential for boundless knowledge sharing in online forums, managers often jump to the conclusion that the potential turned into reality. For instance, based upon their large and diverse pool of participants, managers claim that online forums accomplished knowledge flow across boundaries:

> "Just 18 months after its introduction, 6,000 ShareNet users were registered in 48 countries, and today there are around 16,500 users in more than 70 countries… ShareNet networks these experts globally and lets them share and develop their knowledge in order to create better customer solutions."
>
> *Joachim Döring, manager at Information and Communication Networks Division in Siemens, which deployed an online knowledge forum named 'ShareNet' (Saphorster 2004)*

Systematic studies that investigate whether employees who use these online forums are actually crossing boundaries or staying within them are lacking.

Our research aims to fill the gap on the use of online forums for knowledge sharing. We apply theories of common ground, experience-based learning, and relational demography to examine whether employees transfer knowledge to others within or across boundaries and how the tendencies change as a function of a knowledge provider's experience in an online forum. We propose that participants prefer to transfer knowledge to others with whom they share common ground and use joint characteristics with their communicating partners to assess the level of common ground. In particular, we examine the effects of surface-level (location and status) and deep-level (expertise) similarities of a knowledge source and a knowledge seeker on the likelihood of knowledge transfer. We suggest that both similarities of a potential knowledge-sharing pair drive knowledge transfer. Further we propose that the deep-level similarity effect on knowledge transfer increases whereas the surface-level similarity effect decreases as a knowledge source accumulates more knowledge-sharing experience in an online forum.

We test our hypotheses using panel data in an online forum at a global IT consulting company. We find that both surface- and deep-level similarities of employee pairs increase the likelihood of knowledge transfer for the pairs. That is, employees tend to post answers to questions posted by others who are in the same city, in the same hierarchical status, or who have similar expertise. More interestingly, we find that the effect of deep-level similarity increases whereas the effect of surface-level similarity decreases with more knowledge-sharing experience of a knowledge provider. As participants gain experience sharing knowledge in an online forum, they provide answers less frequently to others at the same location and with the same status but more frequently to others with similar expertise. Our findings suggest that, in an online knowledge community, it can take significant time for participants to find the "right" knowledge-sharing partner because information on deep-level individual characteristics, which allows more accurate assessment of common ground, can only be obtained through extended interaction.

THEORY AND HYPOTHESES

Challenge of Transferring Knowledge in Online

Thanks to an online knowledge forum, scattered employees can now ask for help to a broader audience without any barrier of time and space. Whenever nearby colleagues cannot provide solutions, an employee can enter an online forum to ask for advice. Then, any employee, even strangers, can offer answers to the question poster. One might be skeptical about the quality of advice from strangers.

However, previous research has found that, at least in an organizational context, strangers provide useful information that actually helped seekers to solve problems (Constant et al. 1996).

Although it might sound easy, in fact, transferring knowledge in an online forum is fraught with challenges. In order for communication to be effective, communicating partners should stand on common ground. Common ground is knowledge that communicators share in common and know they share (Clark and Marshall 1981, Krauss and Fussell 1990). It facilitates communication by allowing message senders to customize messages according to message receivers' background knowledge. Without common ground, it is not only hard for communicators to proceed but also the communication is more likely to be misunderstood. The more accurate the assessment of common ground by communicating partners, the more effective that they will be in tailoring their messages and the more successful (understood as intended) their communication will be (Clark and Marshall 1981).

More often than not, knowledge-sharing partners in an online forum do not stand on common ground. Because employees usually share knowledge with strangers in an online forum, they do not have pre-established common ground. Moreover, intermittent interaction with the same partner and limited availability of interactive feedback (i.e., nodding, expression such as "uhuh") are insufficient to find common ground instantaneously. Hence, it can be hard to transfer knowledge in an online forum. For instance, a knowledge provider may easily misinterpret questions. Unaware about what their audience does and does not know, knowledge seekers often do not provide complete information on the problem they are facing. Instead, they skip explaining details assuming that others know the details. Given incomplete information, a knowledge provider is likely to formulate answers using incorrect assumptions based on his or her own knowledge. Moreover, unaware of what a knowledge seeker knows, a knowledge provider may use terminology that the corresponding seeker does not know. What makes things worse online is that any confusion due to the lack of common ground remains unresolved because of the asynchronous nature of communication.

Assessing Common Ground from Similarity in an Online Knowledge Communities

Due to the difficulties of sharing knowledge without common ground, we propose that employees prefer to transfer knowledge to others with common ground, so that their efforts to help others are not wasted. Possessing similar user profile information is likely to signal a higher level of common ground. Even though much information on communicating partners (e.g. voice tone, appearance) disappears in an online forum, some information remains accessible. For example, most web forums make basic user profile information public. The richness of available profile information depends on the policy of an online community and user preference but online forums within an organization tend to provide richer information due to fewer privacy concerns than public online forums. Category membership information such as gender, status, and location are frequently listed items.

An employee may feel that he or she has common ground with other employees if they have similar user profile information. People tend to presume others' background knowledge based on category membership (Clark and Marshall 1981, Krauss and Fussell 1990, McPherson et al. 2001). Thus, knowledge providers may feel that knowledge seekers possess similar knowledge if the seekers are in the same status, location or age. As a result, the knowledge provider may feel more certain that he or she understood a question and more likely to provide a solution to others with similar user profile information than to others with dissimilar user profiles. Also, sharing user profiles can actually provide common ground because people who have similar demographic (e.g., sex, location) or social characteristics (e.g., education, social category membership) are more likely to go through similar experiences, which leads to more knowledge in common (Reagans and McEvily 2003).

In addition to user profile data, employees can obtain information about other employees as they interact with others in an online forum. Because individuals are more likely to be knowledgeable in areas where they provide answers (Zhang et al. 2007), employees can acquire information about others' expertise in the course of asking and providing knowledge in an online forum. Once employees obtain expertise information, they are expected to transfer knowledge more frequently to others with similar rather than dissimilar expertise.

Possessing similar expertise lowers a knowledge source's effort of transfer. Individuals find it easy to transfer knowledge to recipients who have similar background knowledge because a message sender can use prior knowledge as a building block of new knowledge. For instance, individuals with similar expertise are more likely to have a shared language and skills, which eases the knowledge transfer process, than individuals with dissimilar expertise. As an example, compare two cases where a rocket scientist tries to explain the technical reasons why the Space Shuttle Challenger broke apart to another rocket scientist versus to an accountant. The task would be much more strenuous when the knowledge recipient is an accountant than when he or she is a rocket scientist. Simply figuring out the accountant's background knowledge on rocket engineering would take non-trivial time and effort. Additionally it would be challenging for the scientist to rephrase all engineering terms so that the accountant could understand. Similar difficulties would arise when the accountant attempts to explain public audit procedures to the scientist. Moreover, knowledge transfer is more likely to be successful when knowledge recipients possess prior-related knowledge. People find it easier to learn new knowledge in areas where they have prior expertise because they can associate new ideas to pre-existing knowledge (Bower and Hilgard 1981).

Based on the above arguments, we expect that employees will prefer to transfer knowledge to others with similar attributes. Therefore, we hypothesize that:

> *Hypothesis 1. The similarity of a knowledge source and a knowledge seeker increases the likelihood of knowledge transfer.*

Shifting from Surface- to Deep-level Common Ground with Experience

Scholars in relational demography literature characterize individual attributes either as surface-level or deep-level (Harrison et al. 1998, Jackson et al. 1995, Williams and O'Reilly 1998). Surface-level attributes are overt characteristics of individuals that are readily detectable, generally immutable, and easily measurable features. Demographic and social characteristics such as age, gender, and group membership are examples of surface-level attributes. On the other hand, deep-level attributes are covert characteristics of individuals that are "more subject to construal and mutable (Jackson et al. 1995; 217)" than surface-level attributes. Extended interactions and information-gathering processes are generally required to learn about deep-level characteristics of others. Examples of deep-level attributes are attitudes, values, knowledge, and skills.

People tend to form their initial perceptions of others based on surface-level features and adjust those perceptions only after they gain deep-level information about others (Amir 1976, Byrne and Wong 1962, Harrison et al 1998, Stangor et al. 1992). Likewise, in an online forum, we expect that participants make initial assessments of common ground using surface-level information and then replace the superficial, and often inaccurate, initial assessments with more accurate ones based on deep-level information. In an online forum, employees cannot physically see other participants. For that reason, we consider that generally immutable, and easily measureable attributes of employees that are disclosed in their user profiles as surface-level information in the online forum. When new to an online forum, employees are likely to over-weight the user profile information to assess common ground with others because it is the only information available.

From experiences of requesting and providing knowledge in an online forum, employees learn to discern each other's expertise, a deep-level attribute of others. Participants can detect other's expertise by either observing or by directly interacting with others because people are more likely to be knowledgeable on topics where they provide answers. As participants gain more experience using an online forum, we expect them to replace initial assessments of common ground based on surface-level user profile information with the ones based on deep-level expertise information.

In sum, we expect that the relative impact of surface-level versus deep-level similarities on the likelihood of knowledge transfer will be moderated by the cumulative knowledge-sharing experience of a knowledge provider. That is, as employees gain experience sharing knowledge in an online forum, the basis of assessing common ground changes from surface-level information to deep-level information. Therefore, we hypothesize that:

Hypothesis 2a. The effect of deep-level similarity on the likelihood knowledge transfer becomes stronger as a knowledge source's experience in an online knowledge forum increases.

Hypothesis 2b. The effect of surface-level similarity on the likelihood of knowledge transfer becomes weaker as a knowledge source's experience in an online knowledge forum increases.

METHODS

We tested our hypotheses at an online forum in a Fortune 1,000 company ('the organization' hereafter). The organization is a leading information-technology consulting company with over 118,000 employees and revenues of $4.59 billion as of December 31, 2010. Project teams of the organization are located in various cities, countries, and continents. In order to aid knowledge sharing across its geographically dispersed employees, the organization set up an online knowledge forum (the online forum) in April 2006.

The online forum allows employees to ask questions and post answers to others' questions using text-based, asynchronous communication. The online forum is only accessible to employees of the organization. In the online forum, all postings are displayed in one webpage that is visible to all participating employees. Nicknames chosen by employees are used as a main identifier of a message poster but a poster's real name, job title, and office location are disclosed as a public user profile. Any employee can access the user profile information by right clicking nick names.

Employees of the organization are actively participating in the online forum. During 17 months period after launching the forum, over 17,000 answers were offered to 33,964 questions. The broad acceptance on the forum is fueled by the following characteristics. Knowledge developed in one project is useful to other projects. Even though each project team might have different clients, employees are performing similar tasks (e.g., migrating data, coding to develop software). Thus, a solution found in one project is likely to be useful to other teams. Also, most employees of the organization are IT professionals, who are proficient in using the online forum.

Data

The organization provided us two sets of data: knowledge-sharing data and user profile information. The knowledge-sharing data cover knowledge-sharing activities during 17 months (from April 2006 to August 2007), starting from the launch of the online forum. Each knowledge-sharing instance provides us information about who posted the message, when it was posted, and which topic the message addressed. Further, using a unique thread identifier, we were able to match a question with its corresponding answers.

In order to estimate the systematic effects of dyadic similarities on the likelihood of knowledge transfer, we needed data that included both pairs who shared knowledge as well as those who did not. Otherwise, our model will produce biased estimators due to sample-selection issue (Greene 2003). Hence, we constructed a panel data set, which comprises pairs who shared knowledge as well as those who did not. The panel data set was constructed in several steps. First, we identified 586 employees who participated in the online forum during the 17-month period. Then, for each question posted, we created all potential pairings of employees using the 586 participating employees. For example, for a question k posted by employee j, 585 pairings of employees are constructed because 585 employees (except for the question poster j) are potential knowledge providers to the question k. Consequently, 19,868,940 dyads were constructed based on 33,964 questions posted during the 17 months.

Measures

The unit of analysis of this study is a dyad (employee pairs). To minimize reverse causality, we constructed our explanatory and outcome variables in different time points. As illustrated in Figure 1, our explanatory and moderating variables were measured using data preceding a question post. Outcome variables were measured using the data after the question post. Descriptions of all variables are summarized in Table 1.

Insert Figure 1 about here

Knowledge transfer The dependent variable of this study is whether employee i transferred knowledge to employee j's question k ($Transfer_{ijk}$). We consider that knowledge is transferred by employee i to j if employee i posted at least one answer to employee j's question k. $Transfer_{ijk}$ is coded as 1 if knowledge is transferred, and 0 otherwise.

Surface-level similarity The independent variables measuring surface-level similarities of potential knowledge–sharing pairs are *location similarity* and *status similarity*. The online forum discloses office locations and job titles of all participating employees in their user profiles. Based on the user profile information, we measured surface-level similarity of employee pairs. *Location similarity* ($LocationSim_{ij}$) is coded as 1 if a knowledge provider (employee i)'s office is located in the same city with his corresponding seeker (employee j)'s office, and 0 otherwise. *Status similarity* ($StatusSim_{ij}$) was measured based on job title information. We first categorized employees into three status groups:

high, middle, and low based on their job titles. Employees who have job titles with words such as chief, manager, senior, principal, chairman, or director were categorized as high-status employees. Job titles with assistant, junior or trainee were categorized as low-hierarchical status. Remaining job titles were classified as middle status. $StatusSim_{ij}$ is coded in the same manner as $LocationSim_{ij}$: 1 if a knowledge provider and a seeker are in the same status level, 0 otherwise. Employees participating in the online forum were located in 73 cities over five countries. Among 19,868,940 pairs we constructed, 6,358,000 pairs (≈ 32%) were working in the same city. Participating employees were relatively homogeneous with respect to hierarchical status. About 70% of the employees in our data belong to the middle status, 22% in high status and 8% in low status. Approximately 87% of our observations were in the same hierarchical status.

Deep-level similarity Another independent variable of this study, which measures deep-level similarity of a knowledge-sharing pair, is *expertise similarity*. Expertise similarity ($ExpertiseSim_{ijk}$), measures the degree of overlap between expertise sets of a knowledge provider and a seeker up to the point when the seeker posts a question k. To measure $ExpertiseSim_{ijk}$, we first generated each employee's expertise profile. Because providing answers to a certain topic area indicates that the knowledge provider has expertise on the subject topic (Zhang et al. 2007), we counted the number of answers each employee provided for each topic classification and constructed all employees' distribution of expertise sets across 114 topic areas. The 114 topic classifications are pre-specified items in the online forum. Whenever an employee posts a question, he or she should select one topic item that best matches the question topic area. Corresponding answers to the question automatically carry the same topic item. An employee's expertise distribution is captured by a multidimensional vector, $E_{ik} = \left(E_{ik}^{1} \cdots E_{ik}^{S}\right)$, where E_{ik}^{S} represents the number of response postings in expertise area S by employee *i* until question k is posted. A simple example of an expertise profile where there are five expertise categories would be (2 0 0 0 19), which indicates that the employee provided two answers on the first topic item, 19 answers on the fifth topic item and none on the second, third, and fourth topics.

Expertise similarity of two employees is then calculated by using cosine similarity. Expertise similarity determines whether expertise profile vectors of two employees are pointing the same direction by measuring cosign of the angle between the two vectors. Previous studies used cosign similarity to assess the proximity of firms in patent class distribution (Jaffe 1986, Sampson 2007). The formula for *expertise similarity* between a knowledge provider *i* and a seeker *j* for question *k* is defined as:

$$ExpertiseSim_{ijk} \;=\; \frac{E_{ik}E_{jk}^{'}}{\sqrt{\left(E_{ik}E_{ik}^{'}\right)\left(E_{jk}E_{jk}^{'}\right)}}$$

where $i \neq j$. The resulting similarity ranges from 0 to 1: the value of 0 means the two employees do not have any common expertise and 1 means the two have the exact same expertise sets.

Experience The moderating variable of this study, *experience*, is measured by the cumulative number of knowledge-sharing instances of a knowledge provider with any users in the online forum until question *k* posted ($Experience_{ik}$). For each employee, we counted the cumulative number of answer-posting and question-posting instances until question *k* is posted.

Control variables The decision to transfer knowledge can also be driven by other factors. In order to tease out the effects of similarity after controlling for other potential drivers, we incorporated a number of control variables. First, an employee's decision to transfer knowledge may be driven by whether the question poster has provided answers to the employee in the past. To control for reciprocity, we included $Reciprocity_{ijk}$ variable. The measure is calculated by counting the total number of answers provided by employee *j* to employee i's question up to the point when employee *j* posted question k. Second, the decision to transfer knowledge may also be driven by whether a knowledge provider has expertise on the question topic area. We incorporated $Ability_{ik}$ to control for the potential effect of knowledge providers' ability to answer on the likelihood of knowledge transfer. Third, the knowledge-sharing activities span over 17 months. To control for any unobserved effects caused by time differences (e.g., system improvement), our model includes the variable $Time_k$, which indicates the number of weeks passed after the forum launch when question k is posted. Finally, we included random effects for dyads (d_{ij}) as well as employees (a_i, b_j) nested in each dyad to control for any unobserved heterogeneity in the online forum (i.e., activeness in the online forum).

Insert Table 1 about here

Model Specification

Our dyad observations are not independent. Because one employee can be a member of multiple pairs, error terms will be correlated across observations. If we do not account for the dependence, our model will produce artificially reduced standard errors. Among many solutions to the problem (Simpson 2001), we chose to include random effects for each employee in each dyad: a_i for a knowledge provider and b_j for a knowledge seeker. The random effects allow the dependent variable, transfer, to vary randomly around the mean of a dyad across employees within dyads (Greene 2003). Given our dependent variable is dichotomous we use logistic regression to test our hypotheses. The model we estimate takes the general form provided below. Variables are indexed to indicate a knowledge provider (i), a knowledge seeker (j), and a question (k):

$$\begin{aligned} Transfer_{ijk} = f(&LocationSim_{ij}, StatusSim_{ij}, ExpertiseSim_{ijk}, Experience_{ik}, LocationSim_{ij} \\ &* Experience_{ik}, StatusSim_{ij} * Experience_{ik}, ExpertiseSim_{ijk} \\ &* Experience_{ik}, Reciprocity_{ijk}, Ability_{ik}, Time_k, a_i, b_j, d_{ij}) \end{aligned}$$

where the sender (a_i)- and receiver (b_i)-specific effects of the same individual are allowed to be correlated with each other as:

$$\binom{a_i}{b_i} \sim MVN\left(0, \begin{bmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{bmatrix}\right).$$

And the unobserved dyad-specific homophily is captured by using a dyad-specific unobserved random effect, d_{ij}, where $d_{ij} \sim MVN(0, \sigma_d^2)$.[2] Furthermore, we assume that the dyad-specific unobserved effects are symmetric, i.e., $d_{ij} = d_{ji}$.

A final estimation issue concerns with computational challenge. With dataset of 19,868,940 observations, we confronted computational and resource challenges to estimate parameters. This is a challenge often encountered in large-scale dyad-level studies of networks (e.g., Braun and Bonfrer 2009, Lu et al. 2011). To alleviate the challenge, we adopted Bayesian inference procedure for estimation. Because the Bayesian approach does not require maximization algorithms, estimation procedure is more efficient than that of the frequentist approach (Cameron and Trivedi 1998, Gelman and Hill 2007). Contrary to the frequentist approach where the main interest is to determine the point estimate of true parameter value θ_0, the interest of Bayesian approach lies in producing the entire distribution of the parameters of interest given the data and a prior. We estimated the parameters by using a Markov Chain Monte Carlo (MCMC) procedure, using a Gibbs sampler and the Metropolis-Hastings algorithm. Models were estimated with Matlab and the full estimation procedure is provided in Appendix.

RESULTS

The means, standard deviations, and correlations for all variables are reported in Table 2. Table 3 presents the result of panel logistic regression analysis with two-way random effects. Effects are introduced across columns to demonstrate stability of results. Model 1 includes only control variables. Estimates for surface-level similarity, $LocationSim_{ijk}$ and $StatusSim_{ijk}$, are added in model 2 and 3, respectively. $ExpertiseSim_{ijk}$, our estimate for deep-level similarity, is further incorporated in model 4. Models 5 through 8 add interaction terms between $Experience_{ik}$ and the three similarities. Because the

[2] A richer approach for capturing unobserved homophily is to cluster individuals in multi-dimensional space representing latent characteristics. See Braun and Bonfrer (2010) for an excellent application.

direction and significance of all coefficients are stable across models, we use the complete specification (model 8) to discuss the results.

Insert Table 2 about here

Insert Table 3 about here

The results support the first hypothesis, which predicted positive effects of all dyadic similarities on the likelihood of knowledge transfer. The coefficients of all three similarities ($LocationSim_{ij}$, $StatusSim_{ij}$, and $ExpertiseSim_{ijk}$) are positive and statistically significant, indicating that employees choose similar others to transfer knowledge. Our second set of hypotheses is also supported. The interaction terms between surface-level similarities and experience are negative and statistically significant whereas the interaction between deep-level similarity and experience is positive and significant.

Because our interaction terms are statistically significant, the main effects of similarities should be interpreted in conjunction with their interaction effects with $Experience_{ik}$ (Jaccard and Turrisi 2003). The magnitude of similarity effects on the likelihood of knowledge transfer changes according to the level of our moderating variable, knowledge provider's cumulative knowledge-sharing experience in the online forum. Our results demonstrate that the total effect of surface-level similarities on the likelihood of transfer *decreases* whereas the total effect of deep-level similarity *increases,* as knowledge providers gain more knowledge-sharing experience in the online forum. In other words, compared to a knowledge provider who is new to the online forum, an experienced knowledge provider tends to pay less attention to similarities on surface-level factors but more attention to similarities on deep-level factors when choosing to whom to transfer knowledge.

To better understand the meaning of the interaction effects, we present interaction diagrams in Figure 2. For easier interpretation, we used the log-linear form of logistic model where the log likelihood of knowledge transfer is in linear relationship with the estimated parameters. The "high experience" line shows the slope of the similarity effects on the log likelihood of knowledge transfer for high-experienced knowledge providers (the value of experience is set to the maximum, 792). Similarly, the "low experience" line represents the slope of the similarity effects for low-experienced knowledge providers (the value of experience is set to 1). In Figures 2(a) and 2(b), which illustrate the interaction effects between surface-level similarities and experience, the magnitude of similarity effects become smaller as a knowledge provider becomes more experienced. The positive effect of co-location (co-status) becomes neutral (total effect is zero) once a knowledge provider accumulates total knowledge-sharing experience of 2,214.5. Because the interaction effects between status similarity and experience (β = -0.52) is large compared to the main effect of status similarity (β = 0.64), we see that the total effect of status similarity turns to negative for a high-experienced knowledge provider. Conversely, the effects of our deep-level similarity, $ExpertiseSim_{ijk}$, become stronger as one become more experienced, as illustrated in Figure 2(c).

Insert Figure 2 about here

When incorporating interaction terms into the analysis model, multicollinearity could be a concern. If substantial collinearity problem presents, parameter estimates can be unstable to very small changes in the data (Greene 2003). To examine the influence of multicollinearity, we employed a number of robustness checks. First, we computed variance inflation factors (VIF) for each of the independent variables and interactions among them. VIF quantifies the severity of multicollinearity by measuring how much of the variance of coefficients is increased because of correlation among the explanatory variables

(Marquardt 1970). VIF of 5 or 10 and above is suggested to indicate a multicollinearity problem (O'Brien 2007). The test revealed no diagnostic problem in our case: all of our VIF statistics are below 5 with mean VIF across all variables of 1.92. Second, we introduced effects across columns in Table 3 in order to see whether the size or signs of effects changes significantly. The directions and magnitude of effects are consistent across columns, demonstrating the robustness of results. Another factor that gives more credibility on our result is that our large data produces a high level of statistical power. It has been found that large data can overcome even extremely high correlations among variables (Mason and Perreault 1991).

DISCUSSION AND CONCLUSIONS

Information technology (IT) opened up vast opportunities for companies to make the most out of existing knowledge base. The great opportunity comes from the '*theoretical'* potential of IT to connect otherwise unconnected people, unconstrained by boundaries. However, this great opportunity might be an illusion because IT, paradoxically, also has a potential to fragment communities (Van Alstyne and Brynjolfsson 1995). For instance, people might use advanced filtering capability to locate the most like-minded others and interact only with them, causing boundaries to shift rather than vanish (Van Alstyne and Brynjolfsson 1995).

Motivated by the puzzle, we proposed and empirically tested a theory about how individuals choose communicating partners in an online forum and why. We argued that employees tend to choose similar others who are within boundaries because similar people are more likely to have higher common ground: common ground makes knowledge transfer less effortful but more successful (Clark and Marshall 1981, Krauss and Fussell 1990, McPherson et al. 2001). Further, we claimed that, as employees learn about other users through extended interactions, they favor those who are similar in deep-level attributes as knowledge-sharing partners over those who are similar in surface-level attributes. Consistent with our predictions, the results showed that employees transfer knowledge more frequently to similar others compared to dissimilar others. In addition, as employees gain more knowledge-sharing experience in an online forum, they pay less attention on surface-level similarities (location and hierarchical status) and more attention to deep-level similarity (expertise).

To illustrate how the effects of surface-level similarities on knowledge transfer fade away as a knowledge provider become experienced, we split our observations based on criteria specified in table 4. We found starkly different patterns of each split in terms of how far knowledge providers reach out to offer knowledge (Figure 3). Compared to knowledge providers who are new to the online forum (no experience), high-experienced knowledge providers cross status boundaries about 27 times more on average as depicted in Figure 3A. More interestingly, an answer by high-experienced providers travels about nine times further away, on average, than an answer by new comers: an average answer by a new comer flies 581 miles (approximately the same distance between Chicago downtown to Atlanta) whereas an average answer by high-experienced knowledge provider travels 5,182 miles (approximately the same distance between San Jose in California, U.S.A. to London in United Kingdom) (Figure 3B).

Insert Table 4 about here

Insert Figure 3 about here

In contrast to fading surface-level boundaries, deep-level boundary strengthens as a knowledge provider accumulates experience. The average expertise similarities of knowledge-sharing pairs whose providers are new, low-experienced, and high-experienced to the forum are 0.4, 0.57, and 0.8 respectively. The increasing average expertise similarity value suggests that knowledge providers are getting better in finding out seekers who have similar background knowledge with them.

Several steps were taken to rule out alternative interpretations of our findings. First, the observed effects could be due to different opportunities. For example, if participants in the online forum became more diverse in terms of their surface-level attributes, a knowledge provider will have more opportunity to share knowledge with employees who are dissimilar on surface. Similarly, if the number of topic areas asked in the online forum shrinks over time, a knowledge provider will have less opportunity to share diverse types of knowledge, leading to positive effects of the interaction between experience and expertise similarity. As a robustness check, we measured the change in diversity of participants' demographic attributes and topic areas in the online forum over 16 months. We employed an entropy-based index, Shannon's diversity index, which measures the richness and evenness of different categories inside a population. We found that diversity of the forum participants' office locations, hierarchical status and the number of topic areas discussed were stable during the 16 months.

Second, in order to tease out the effects of similarity after controlling for other potential drivers of knowledge transfer, we incorporated a number of control variables: $Reciprocity_{ijk}$, $Ability_{ik}$, $Time_{k}$. In addition, we included random effects for each employee to control for any unobserved heterogeneity of employees. The results remain consistent after incorporating the control variables.

Our study makes several important contributions. First, to the best of our knowledge, ours is the first empirical study with large panel data that examined whether IT unites or fragments communities. Scholars have shown significant interests in online communities but most of them have focused on individual-level motivation to contribute (e.g., Bagozzi and Dholakia 2006, Ma and Agarwal 2007, Ren et al. 2007, von Hippel and von Krough 2003, Wasko and Faraj 2005). Even though the individual-level studies provide valuable insight, our dyad-level approach offers complementary insights on with whom individuals interact online and why.

Second, our results provide empirical evidence in support of Van Alstyne and Brynjolfsson (2005)'s suggestion that IT shifts boundaries from geography to interests. In spite of the global connectivity from the Internet, our findings suggest that an online knowledge community is fragmented by several boundaries. Yet the surface-level boundaries (e.g., geography, hierarchical status) can be supplanted by 'more desirable' deep-level boundaries with significant knowledge-sharing experience of participants. Our findings are also consistent with the view that online communities are fluid objects that are constantly morphing their boundaries, participants, and interactions (Faraj et al. 2011).

Third, although researchers have argued that "irrelevant" cues are filtered out in computer-mediated communication (Kiesler et al. 1984, Lancaster 1978, Linstone and Turoff 1975, Sproull and Kiesler 1986), our findings indicate that participants in online forums attend to those cues. Our finding that participants use surface-level information to approximate common ground with others suggests that boundaries along surface-level attributes continue to exist online. Although surface-level information may be "irrelevant" cues in online, individuals apply the information to make a rough guess about others until they learn more "relevant" deep-level information through extensive interactions with others.

Finally, we discuss the novelty of solutions exchanged in an online knowledge community. The result suggests that a strong driver of online interactions is homophily. The homophily principle means that contact between similar people occurs at a higher rate than among dissimilar people (McPherson et al. 2001). Because similar people are more likely to know similar things, solutions obtained from an online forum are likely to be similar to the existing knowledge base. Therefore, an online forum seems to foster exploitative rather than exploratory search in organizational learning (March 1991).

This study has a few limitations that suggest directions for future research. First, the current study focused on behaviors of active participants (who contribute to the online forum by posting questions or answers) because our interest was to find out how individuals choose to whom to transfer knowledge and why. Although the behavior of inactive users was not the focus of our current study, exploring how inactive participants (ones who can use community resources but do not contribute) use community resources will provide valuable insight on how knowledge is distributed through an online forum. Another limitation is that our results might not be generalizable to web forums where discussion topics are more diverse. Public web forums such as Yahoo! Answers let users discuss various topics including marriage, sports, travel, and health. The patterns of user interactions were found to be different for different topics (Adamic et al. 2008). Theorizing and empirically testing the relationship between motivations of participants for different topic categories and the resulting patterns of user interactions will advance our knowledge on how individuals form relationships in online.

Managers can utilize our findings to boost boundary-spanning knowledge transfer in online forums. First, incentives to entice employees to come back to an online forum more often will help them to pick up deep-level information and thereby pay less attention to surface-level similarity. Second, managers can make deep-level information more salient and surface-level information less salient so that less time is required for employees to pick up the "more relevant" deep-level information.

In conclusion, this study applies theories of experience-based learning, common ground, and relational demography to investigate how people share knowledge in online communities. We find that people prefer to transfer knowledge to similar others. Additionally, we find that as people learn about other users through extended interactions, they favor those who are similar in deep-level attributes as a knowledge-sharing partner over those who are similar in surface-level attributes. Surface-level boundaries weaken whereas boundaries along deep-level attributes strengthen as participants accumulate more knowledge-sharing experience in an online forum.

REFERENCES

Adamic, L. A, J. Zhang, E. Bakshy, and M. S Ackerman. 2008. "Knowledge sharing and yahoo answers: everyone knows something." Pp. 665–674 in *Proceeding of the 17th international conference on World Wide Web*.

Van Alstyne, M. W, and E. Brynjolfsson. 2005. "Global village or CyberBalkans: Modeling and measuring the integration of electronic communities." *Management Science* 51(6):851-868.

Amir, Y. 1969. "Contact hypothesis in ethnic relations." *Psychological Bulletin* 71(5):319.

Argote, L. 1999. *Organizational learning: Creating, retaining, and transferring knowledge*. Springer Netherlands.

Bagozzi, R.P., and U.M. Dholakia. 2006. "Open source software user communities: A study of participation in Linux user groups." *Management Science* 52(7):1099-1115.

Bower, G.H., and E.R. Hilgard. 1981. *Theories of learning*. NJ: Prentice-Hall.

Braun, M., and A. Bonfrer. 2009. *Censoring, Interdependence and Scalability for Dyadic Social Media Data*. Massachusetts Institute of Technology.

Byrne, D., and T.J. Wong. 1962. "Racial prejudice, interpersonal attraction, and assumed dissimilarity of attitudes." *The Journal of Abnormal and Social Psychology* 65(4):246.

Cameron, AC, and PK Trivedi. 1998. *Microeconometrics: Methods and Applications*. Cambridge University Press.

Clark, H., and C. Marshall. 1981. "Definite reference and mutual knowledge." Pp. 10-63 in *Elements of Discourse Understanding*. New York: Cambridge University Press.

Cohen, W. M, and D. A Levinthal. 1990. "Absorptive capacity: a new perspective on learning and innovation." *Administrative Science Quarterly* 35(1).

Constant, D., L. Sproull, and S. Kiesler. 1996. "The kindness of strangers: The usefulness of electronic weak ties for technical advice." *Organization Science* 7(2):119–135.

Donath, J. S. 1995. “Visual Who: Animating the affinities and activities of an electronic community.” P. 107 in *Proceedings of the third ACM international conference on Multimedia*.

Faraj, Samer, Sirkka Javenpaa, and Ann Majchrzak. 2011. “Knowledge collaboration in online communities.” *Organization Science*.

Framington, H. 2007. “CIOs use WEB 2.0 to keep up with competition: study.” *Computer World*.

Gelman, A., and J. Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press New York.

Grant, R. M. 1996. “Toward a knowledge-based theory of the firm.” *Strategic Management Journal* 17(10):109–122.

Greene, William. 2003. *Econometric analysis*. 5th ed. Englewood Cliffs, NJ: Prentice Hall.

Harrison, David, Kenneth Price, and Myrtle Bell. 1998. “Beyond relational demography: Time and the effects of surface- and deep-level diversity on work group cohesion.” *Academy of Management Journal* 41(1):96-107.

Von Hippel, E., and E. A Von Hippel. 1988. *The sources of innovation*. Oxford University Press New York.

Jaccard, J., and R Turrisi. 2003. *Interaction effects in multiple regression*. Thousand oaks, CA: Sage Publications, Inc.

Jackson, May, and Whitney. 1995. “Understanding the dynamics of diversity in decision-making teams.” Pp. 204-261 in *Team decision-making effectiveness in organizations*. R.A.Guzzo & E.Salas (Eds.).

Jaffe, A. B. 1986. “Technological opportunity and spillovers of R & D: evidence from firms’ patents, profits, and market value.” *American Economic Review* 76(5):984–1001.

Kiesler, S., J. Siegel, and T. W McGuire. 1984. “Social psychological aspects of computer-mediated communication.” *American Psychologist* 39(10):1123.

Kietzmann, J.H., K. Hermkens, I.P. McCarthy, and B.S. Silvestre. 2011. “Social media? Get serious! Understanding the functional building blocks of social media.” *Business Horizons*.

Krauss, R.M., and S.R. Fussell. 1990. "Mutual knowledge and communicative effectiveness, Intellectual teamwork: social and technological foundations of cooperative work." Pp. 111-146 in *Intellectual teamwork: Social and technological foundations of cooperative work*. Hillsdate, NJ: Lawrence Erlbaum Associates, Inc.

Lancaster, F.W. 1978. *Toward paperless information systems*. Academic Press, Inc.

Linstone, H.A., M. Turoff, and others. 1976. *The Delphi method: Techniques and applications*. Addison-Wesley.

Lu, Yingda, Kinshuk Jerath, and Vir Singh Param. 2011. "The Emergence of Opinion Leaders in Online Review Communities." *working paper*.

Ma, Meng, and Ritu Agarwal. 2007. "Through a glass darkly: Information technology design, identity verification, and knkowledge contribution in online communities." *Information Systems Research* 18(1):42-67.

March, J. G. 1991. "Exploration and exploitation in organizational learning." *Organization Science* 2(1):71–87.

Marquardt, D.W. 1970. "Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation." *Technometrics* 591–612.

Mason, C.H., and W.D. Perreault. 1991. "Collinearity, power, and interpretation of multiple regression analysis." *Journal of Marketing Research* 268–280.

McPherson, M., L. Smith-Lovin, and J. M Cook. 2001. "Birds of a feather: Homophily in social networks." *Annual Review of Sociology* 27:415-444.

O'Brien, Robert. 2007. "A caution regarding rules of thumb for variance inflation factors." *Quality and Quantity* 41(5):673-690. Retrieved August 26, 2010.

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

Ren, Yuqing, R. Kraut, and S. Kiesler. 2007. "Applying Common Identity and Bond Theory to Design of Online Communities." *Organization Studies* 28(3):377-408. Retrieved August 31, 2011.

Sampson, R. C. 2007. “R&D alliances and firm performance: the impact of technological diversity and alliance organization on innovation.” *Academy of Management Journal* 50(2):364.

Saphorster, S. 2004. “Heading for Knowledge-Guided Networks.” *Pictures of the Future*.

Simpson, William. 2001. “The Quadratic Assignment Procedure (QAP).”

Sproull, L., and S. Kiesler. 1991. “New ways of working in the networked organization.” *Cambridge, MA*.

Sproull, L., and S. Kiesler. 1986. “Reducing social context cues: Electronic mail in organizational communications.” *Management Science* 32(11):1492–1512.

Stangor, C., L. Lynch, C. Duan, and B. Glas. 1992. “Categorization of individuals on the basis of multiple social features.” *Journal of Personality and Social Psychology* 62(2):207.

Wasko, M.M.L., and S. Faraj. 2005. “Why should I share? Examining social capital and knowledge contribution in electronic networks of practice.” *Mis Quarterly* 35–57.

Williams, Katherine, and Charles O’Reilly. 1998. “Demography and diversity in organizations: A review of 40 years of research.” *Research in Organizational Behavior* 20:77–140.

Zander, U., and B. Kogut. 1995. “Knowledge and the speed of the transfer and imitation of organizational capabilities: An empirical test.” *Organization Science* 6(1):76–92.

Zhang, J., M. S Ackerman, and L. Adamic. 2007. “Expertise networks in online communities: structure and algorithms.” Pp. 221–230 in *Proceedings of the 16th international conference on World Wide Web*.

FIGURE 1

Illustration of Panel Dataset Construction

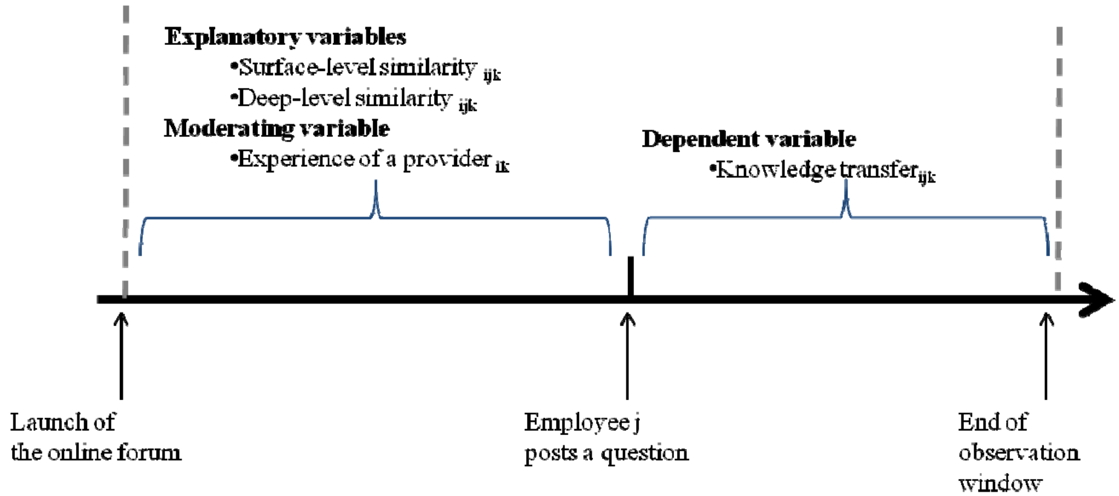


TABLE 1

Descriptions of Variables

Variables	Definition
Dependent variable	
$Transfer_{ijk}$	Binary variable indicating whether employee *i* posted an answer to employee j's question k
Independent variables	
Surface-level similarity	
$LocationSim_{ij}$	Binary variable indicating whether offices of employee *i* and employee *j* are in the same city : 1 if in the same city, 0 otherwise
$StatusSim_{ij}$	Binary variable indicating whether employee *i* and employee *j* are in the same hierarchical status : 1 if in the same hierarchical status, 0 otherwise
Deep-level similarity	
$ExpertiseSim_{ijk}$	The degree of overlap between expertise profiles of employee *i* and *j* at the point when *j* posted question *k*
Moderating variable	
$Experience_{ik}$	The sum of cumulative number of a knowledge provider (employee i)'s answer-posting and question-posting instances until question *k* is posted
Control variables	
$Reciprocity_{ijk}$	Number of answers provided by employee *j* to employee *i's* question up to the point when question *k* is posted by *j*
$Ability_{ik}$	Binary variable indicating whether employee *i* have expertise on the topic area of question *k* : 1 if yes, 0 otherwise
$Time_k$	The number of weeks passed after the online forum launch when question k is posted

TABLE 2

Descriptive Statistics and Correlations

	Mean	s.d.	1	2	3	4	5	6	7	8
1 $LocationSim_{ij}$	0.317	0.455	1.000	0.004	0.016	0.063	0.018	0.101	0.022	0.023
2 $StatusSim_{ij}$	0.871	0.325		1.000	0.059	0.114	0.093	0.066	0.127	0.010
3 $ExpertiseSim_{ijk}$	0.116	0.290			1.000	0.105	-0.202	0.052	0.201	0.014
4 $Experience_{ik}$	19.866	1.196				1.000	0.131	0.216	0.281	0.054
5 Time	13.464	4.295					1.000	0.190	0.161	-0.005
6 $Reciprocity_{ijk}$	0.143	0.365						1.000	0.109	0.045
7 $Ability_{ik}$	0.016	0.372							1.000	0.025
8 $Transfer_{ijk}$	0.002	0.049								1.000

N=19,868,940

TABLE 3

Predicting Knowledge Transfer: Panel Logistic Regression with Two-way Random Effects

	1	2	3	4	5	6	7	8
Surface-level similarity								
$LocationSim_{ij}$		0.89 ***	0.89 ***	0.68 ***	0.94 ***	1.15 ***	1.55 ***	1.55 ***
$StatusSim_{ij}$			0.95 ***	0.43 ***	0.67 ***	0.43 ***	0.64 ***	0.64 ***
Deep-level similarity								
$ExpertiseSim_{ijk}$				0.27 ***	0.29 ***	0.41 ***	0.40 ***	0.29 ***
Experience								
$Experience_{ik}$					0.49 ***	0.51 ***	0.81 ***	0.70 ***
Interactions								
$LocationSim_{ij}$ x $Experience_{ik}$						-0.20 ***	-0.21 ***	-0.20 ***
$StatusSim_{ij}$ x $Experience_{ik}$							-0.52 ***	-0.52 ***
$ExpertiseSim_{ijk}$ x $Experience_{ik}$								0.40 ***
Control								
$Reciprocity_{ijk}$	0.44 ***	0.42 ***	0.40 ***	0.19 ***	0.24 **	0.24 **	0.24 **	0.25 **
$Ability_{ik}$	1.42 ***	0.98 ***	0.83 ***	0.72 **	0.64 **	0.61 **	0.61 **	0.62 **
Time	-0.41 *	-0.56 **	-0.68 **	-0.38 *	-1.06 ***	-1.16 ***	-1.16 ***	-1.43 ***
Random effects								
Provider mean effect	-1.90 ***	-1.90 ***	-1.90 ***	-1.90 ***	-1.90 ***	-1.90 ***	-1.90 ***	-1.90 ***
Seeker mean effect	-4.79 ***	-4.79 ***	-4.79 ***	-4.79 ***	-4.79 ***	-4.79 ***	-4.79 ***	-4.79 ***

Dependent variable: $Transfer_{ijk}$

N = 19,868,940

**** The 99% credible interval does not include zero.*

*** The 95% credible interval does not include zero.*

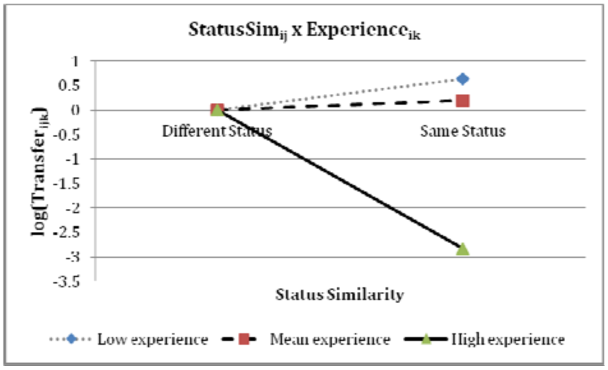
** The 90% credible interval does not include zero.*

FIGURE 2

Interaction Diagrams

(a) Location similarity x Experience

(b) Status similarity x Experience



(c) Expertise similarity x Experience

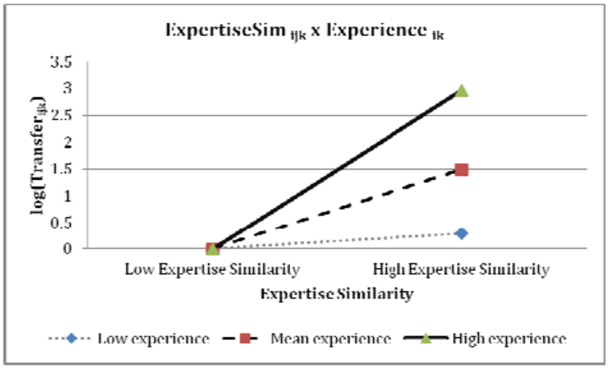


TABLE 4

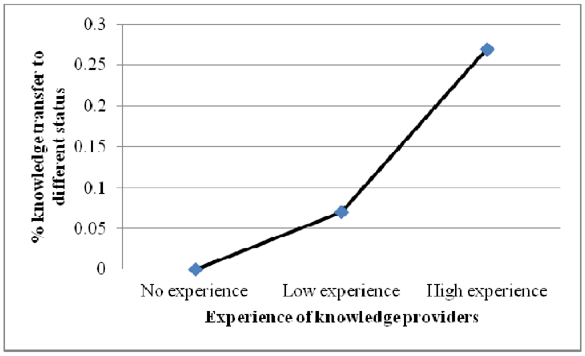
Sample Split Criteria

Experience of knowledge providers	Split rule	% of Dyads
No experience	Experience=0	5.1%
Low experience	0 < Experience ≤ median experience	44.9%
High experience	Experience > median experience	50.0%

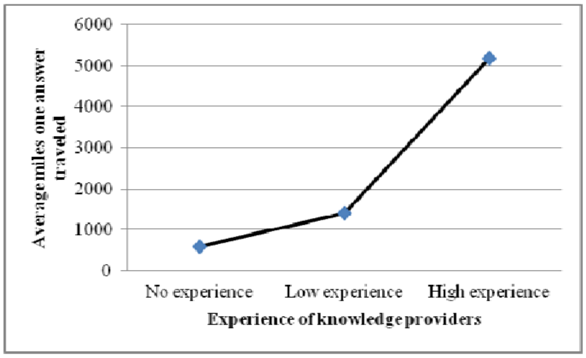
FIGURE 3

Fading Surface-level Boundaries with Experience

(A) % of answers to employees with different status



(B) Distance knowledge traveled



Experience of knowledge providers	% Knowledge transfer to different status employees	x No experience
No experience	1.0%	1x
Low experience	7.0%	7x
High experience	27.0%	27x

Experience of knowledge providers	Average miles knowledge traveled	x No experience
No experience	581.0	1.0x
Low experience	1,393.4	2.4x
High experience	5,182.1	8.9x

APPENDIX

Parameter Estimation Procedure

For the procedures below, letters with superscript u represent the values of the updated corresponding parameters.

Step 1: Estimating $\boldsymbol{\gamma}$ **(**$\boldsymbol{\gamma}$ represents homogeneous coefficients**)**

$$\boldsymbol{\gamma}^u | a_i, b_i, \alpha_0, \alpha_1, d_{ij}, data$$

$$f\left(\boldsymbol{\gamma}^u \middle| a_i, b_i, \alpha_0, \alpha_1, d_{ij}, data\right)$$

$$\propto \left|\Sigma_{\boldsymbol{\gamma}\mathbf{0}}\right|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\boldsymbol{\gamma}^u - \overline{\boldsymbol{\gamma}_\mathbf{0}})' \Sigma_{\boldsymbol{\gamma}0}^{-1} (\boldsymbol{\gamma}^u - \overline{\boldsymbol{\gamma}_\mathbf{0}})\right] L(\boldsymbol{Y})$$

where $\overline{\boldsymbol{\gamma}_\mathbf{0}}$ and $\Sigma_{\boldsymbol{\gamma}\mathbf{0}}$ are diffused priors. Because there is no closed form for this, we use the Metropolis-Hastings algorithm to draw from this conditional distribution of $\boldsymbol{\gamma}^u$. The probability of accepting $\boldsymbol{\gamma}^u$ is:

$$\Pr(acceptance) = \min\left\{\frac{\exp\left[-\frac{1}{2}(\boldsymbol{\gamma}^u - \overline{\boldsymbol{\gamma}_\mathbf{0}})' \Sigma_{\boldsymbol{\gamma}0}^{-1} (\boldsymbol{\gamma}^u - \overline{\boldsymbol{\gamma}_\mathbf{0}})\right] L(\boldsymbol{Y}|\boldsymbol{\gamma}^u)}{\exp\left[-\frac{1}{2}(\boldsymbol{\gamma} - \overline{\boldsymbol{\gamma}_\mathbf{0}})' \Sigma_{\boldsymbol{\gamma}0}^{-1} (\boldsymbol{\gamma} - \overline{\boldsymbol{\gamma}_\mathbf{0}})\right] L(\boldsymbol{Y}|\boldsymbol{\gamma})}, 1\right\}$$

We define diffuse priors by setting $\overline{\boldsymbol{\gamma}_\mathbf{0}}$ to be a vector of zeros and $\Sigma_{\boldsymbol{\gamma}\mathbf{0}} = 30I$.[3]

Step 2: Generate a_i^u, b_i^u :

$$f\left(a_i^u, b_i^u | \boldsymbol{\gamma}^u, \alpha_0^u, \alpha_1^u, d_{ij}, data\right)$$

$$\propto N\left(\left(a_i^u, b_i^u | \boldsymbol{\beta}^u, \alpha_0^u, \alpha_1^u, d_{ij}\right), \Sigma_{ab}\right) L(\boldsymbol{Y})$$

[3] Our estimation is not sensitive to the setting of the diffuse hyperprior.

$$\propto |\Sigma_{ab}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(a_i^u, b_i^u)\Sigma_{ab}^{-1}(a_i^u, b_i^u)'\right] L(\boldsymbol{Y})$$

Because this distribution does not have a closed form, we use the Metropolis-Hastings algorithm to draw from the conditional distribution of a_i, b_i: a_i, b_i is the draw of the random effect from the previous iteration, and we draw a_i^u, b_i^u by $\begin{bmatrix} a_i^u \\ b_i^u \end{bmatrix} = \begin{bmatrix} a_i \\ b_i \end{bmatrix} + \Delta \begin{bmatrix} a \\ b \end{bmatrix}$, where $\Delta \begin{bmatrix} a \\ b \end{bmatrix}$ is a draw from $N(0,\Delta^2\Lambda)$, and Δ and Λ are chosen adaptively to reduce autocorrelation among MCMC draws following Atchade (2006). The probability of accepting this $\begin{bmatrix} a_i^u \\ b_i^u \end{bmatrix}$, the updated value for $\begin{bmatrix} a_i \\ b_i \end{bmatrix}$ is:

$$\Pr(acceptance) = \min\left\{\frac{\left[\exp\left(-\frac{1}{2}(a_i^u, b_i^u)\Sigma_{ab}^{-1}(a_i^u, b_i^u)'\right)\right] L(\boldsymbol{Y}|a_i^u, b_i^u)}{\left[\exp\left(-\frac{1}{2}(a_i, b_i)\Sigma_{ab}^{-1}(a_i, b_i)'\right)\right] L(\boldsymbol{Y}|a_i, b_i)}, 1\right\}$$

Step 3: $\Sigma_{\mathrm{ab}}^{\mathrm{u}}|a_i^u, b_i^u$

$$(\Sigma_{\mathrm{ab}}^{\mathrm{u}}|a_i^u, b_i^u) \sim IW_2(7 + N, G_0^{-1} + \sum_{i=1}^{N}(a_i^u, b_i^u)(a_i^u, b_i^u)')$$

Step 4: $d_{ij}^u d_{ij}^u, d_{ji}^u|\alpha_0^u, \boldsymbol{\gamma}^u, a_i, b_i, \alpha_1^u, \sigma_d^2, data$

$$f(d_{ij}^u, d_{ji}^u|\alpha_0^u, \boldsymbol{\gamma}^u, a_i, b_i, \alpha_1^u, \sigma_d^2, data)$$

$$\propto N\left(\left(d_{ij}^u, d_{ji}^u\middle|\alpha_0^u, \boldsymbol{\gamma}^u, a_i, b_i, \alpha_1^u\right), \sigma_d^2\right) L(\boldsymbol{Y})$$

$$\propto \sigma_d^{-1} \exp\left[-\frac{1}{2}\left(d_{ij}^u + d_{ji}^u\right)^2 \sigma_d^{-2}\right] L(\boldsymbol{Y})$$

We use the Metropolis-Hastings algorithm to draw from this conditional distribution of d_{ij}^{u} and d_{ji}^{u}: d_{ij} and d_{ji} are the draw of the unobservable similarity effects from the previous iteration, and we draw d_{ij}^{u}, d_{ji}^{u} by $\begin{bmatrix} d_{ij}^{u} \\ d_{ji}^{u} \end{bmatrix} = \begin{bmatrix} d_{ij} \\ d_{ji} \end{bmatrix} + \Delta\boldsymbol{d}$, where $\Delta\boldsymbol{d}$ is a draw from N(0,$\Delta^2\Lambda$), and Δ and Λ are chosen adaptively to reduce autocorrelation among MCMC draws following Atchade (2006). The probability of accepting $\begin{bmatrix} d_{ij}^{u} \\ d_{ji}^{u} \end{bmatrix}$ is:

$$\Pr(acceptance) = \min\{\frac{\left[\exp\left(-\frac{1}{2}\left(d_{ij}^{u} + d_{ji}^{u}\right)\sigma_d^{-2}\right)\right] L(\boldsymbol{Y}|d_{ij}^{u}, d_{ji}^{u})}{\left[\exp\left(-\frac{1}{2}\left(d_{ij} + d_{ji}\right)\sigma_d^{-2}\right)\right] L(\boldsymbol{Y}|d_{ij}, d_{ji})}, 1\}$$

Step 5: Generating σ_d^{u}

$$(\sigma_d^{u}|\, d_{ij}^{u}, d_{ji}^{u}) \sim IW_1(1 + N(N-1), 1 + \textstyle\sum_{i=1}^{N}\sum_{j=1, j\neq i}^{N}(d_{ij}^{u} + d_{ji}^{u})^2)$$

Step 6: Go to step 1.