Information Systems Research

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

IT Knowledge Spillovers, Absorptive Capacity, and Productivity: Evidence from Enterprise Software

Peng Huang, Marco Ceccagnoli, Chris Forman, D.J. Wu

http://pubsonline.informs.org/journal/isre

INFORMATION SYSTEMS RESEARCH
Vol. 33, No. 3, September 2022, pp. 908–934
ISSN 1047-7047 (print), ISSN 1526-5536 (online)

IT Knowledge Spillovers, Absorptive Capacity, and Productivity: Evidence from Enterprise Software

Peng Huang,[a,*] Marco Ceccagnoli,[b] Chris Forman,[c] D.J. Wu[b]

[a] Robert H. Smith School of Business, University of Maryland, College Park, Maryland 20742; [b] Scheller College of Business, Georgia Institute of Technology, Atlanta, Georgia 30308; [c] Charles H. Dyson School of Applied Economics and Management, Cornell SC Johnson College of Business, Cornell University, Ithaca, New York 14853

*Corresponding author

Contact: huang@umd.edu, https://orcid.org/0000-0002-7872-0396 (PH); marco.ceccagnoli@scheller.gatech.edu (MC); chris.forman@cornell.edu, https://orcid.org/0000-0003-3055-1607 (CF); dj.wu@scheller.gatech.edu, https://orcid.org/0000-0002-7991-1127 (DJW)

Received: August 14, 2017
Revised: September 30, 2019; July 30, 2020; August 2, 2021
Accepted: October 22, 2021
Published Online in Articles in Advance: February 10, 2022

https://doi.org/10.1287/isre.2021.1091

Abstract. We examine the productivity implications of external knowledge flows obtained through an internet-mediated discussion forum in which IT professionals help one another solve problems related to the implementation and use of enterprise software. We extend elements of the absorptive capacity (ACAP) framework that have not previously been studied in the information systems (IS) literature to a new context. Consistent with prior results from the IS literature, we first show that IT spillovers—acquired through employees' participation in this forum—only accrue to firms with prior related investments in enterprise software. We then demonstrate boundary conditions for ACAP based on characteristics of external knowledge affecting the ease of learning. Our results show that IT spillovers are not "free"; the ability to derive the value of IT spillovers through informal channels—such as online communities—critically depends on *both* prior related IT investments by the recipient firm *and* the novelty of external knowledge. Less intuitively, when knowledge originates from relatively novel or emergent domains, the role of prior related knowledge in absorbing spillovers becomes more important.

History: Michael Zhang, Senior Editor; Yuliang (Oliver) Yao, Associate Editor.

Supplemental Material: The online appendix is available at https://doi.org/10.1287/isre.2021.1091.

Keywords: IT spillovers • absorptive capacity • business value of IT • enterprise software • IT human capital

1. Introduction

Whereas information technology (IT) systems are shown to create significant value for the firms that adopt them, the returns often appear with a delay (e.g., Brynjolfsson and Hitt 2003) and may vary greatly across firms (Bresnahan et al. 2002, Aral and Weill 2007, Bloom et al. 2012). Firms investing in new IT systems must often undertake complementary innovation, sometimes termed coinvention, to adapt general-purpose IT systems to the idiosyncratic needs of organizations (Bresnahan and Greenstein 1996). Whereas, sometimes, these innovations are related to technical adaptations to IT hardware and software systems, they also frequently involve changes to organizational elements, such as business processes (Bresnahan et al. 2002, Bartel et al. 2007, Dranove et al. 2014).

The human capital required to deploy these systems is scarce and unequally distributed.[1] However, firms can use many formal and informal means to acquire the necessary knowledge from external sources.[2] The available means range from hiring workers who have acquired the expertise by working on similar projects at other firms (Tambe and Hitt 2014a), knowledge transfer from third-party consultants who have been contracted by the firm (e.g., Ko et al. 2005, Chang and Gurbaxani 2012b), to knowledge exchange that is mediated by communication with industry or supply chain participants (Caselli and Coleman 2001, Chang and Gurbaxani 2012a). The literature on IT spillovers shows that the use of these means can have a significant effect on firm productivity (Cheng and Nault 2007, Chang and Gurbaxani 2012a, Cheng and Nault 2012, Tambe 2014, Tambe and Hitt 2014a). However, recent work calls for a deeper understanding of the mechanisms through which they work (Ba and Nault 2017). In this paper, we aim to respond to this call and deepen our understanding of when IT spillovers to a firm are most beneficial by applying a well-known framework from the R&D literature.

Specifically, an established R&D literature on knowledge spillovers (Cohen and Levinthal 1989, 1990) shows that the effects of knowledge flows on productivity is critically conditioned by a recipient firm's absorptive capacity (ACAP), defined as a firm's ability to assimilate, transform, and apply external knowledge (Cohen and Levinthal 1989, 1990). In the context of IS research, the ACAP concept is conceptualized as the extent of prior IT-related knowledge possessed by an organization (Roberts et al. 2012). Earlier empirical findings suggest that a recipient's ACAP impacts the amount of IT knowledge transferred from a knowledge source and reduces knowledge barriers, thus facilitating IT adoption. For example, within the context of the IT spillover literature, Chang and Gurbaxani (2012a) find that firms that have greater prior IT investments receive higher spillover benefits.

Despite the extensive use of the ACAP concept in IS research, a recent review concludes that "there have been few detailed investigations of the relationship between IT and absorptive capacity" (Roberts et al. 2012, p. 640). In a similar spirit, we argue that the IS literature has overlooked a key exogenous driver of a firm's ACAP as proposed by the original work of Cohen and Levinthal (1989, 1990): the characteristics of outside knowledge that make its learning more difficult. In this paper, we seek to address this limitation by first introducing a model closely aligned with the original model of Cohen and Levinthal (1989), in which prior related investments in IT systems, characteristics of outside IT knowledge, and their interaction influence the process of knowledge accumulation through external sources. We then develop a new measurement strategy for observing knowledge flows related to a firm's IT systems by directly studying the activity of IT workers on an online discussion forum.

Online discussion forums play an increasingly important role in diffusing knowledge within the software development community;[3] for example, the discussion forum Stack Overflow has more than 100 million users,[4] and a recent survey of that site shows that more than 85% of its users visit the site multiple times per week.[5] The SAP Community Network (SCN), which forms the base of our empirical analysis in the context of enterprise software, has an average of two million unique visitors each month.[6] The SCN enables us to track knowledge flows among its users and characteristics of the exchanged knowledge during a five-year period. We then use these measures to estimate a productivity model augmented by a factor of production that captures a firm's knowledge stock, which is critically affected by a firm's ACAP.

We find that external IT knowledge flows are absorbed more readily when firms have made investments in prior related knowledge. This is consistent with prior work that examines the implications of IT knowledge flows on productivity (Chang and Gurbaxani 2012a). However, we extend this work in important ways: we also show that, when external knowledge is more difficult to learn, as when it is novel, more complex, or less targeted to the recipient's needs, a firm's ACAP is lower (Cohen and Levinthal 1989, 1990). Finally, and less intuitively, in a more difficult learning environment, the role of prior related knowledge in building ACAP becomes more important (Cohen and Levinthal 1989, 1990).

We contribute to recent work seeking to understand the productivity benefits of IT investment and related business process innovation in several ways. First, as noted, we advance recent work trying to understand the productivity benefits of external knowledge acquisition or "IT spillovers" (e.g., Chang and Gurbaxani 2012a, b; Tambe and Hitt 2014a) by demonstrating the conditions under which external knowledge acquisition leads to higher productivity. Recent work highlights heterogeneity in the value of IT knowledge spillovers that are mediated by *formal* channels, such as employment relationships (Tambe and Hitt 2014a, Wu et al. 2017). These results are valuable; however, this type of knowledge acquisition is expensive and may not be appropriate for all circumstances.

In contrast, there is less progress in understanding heterogeneity in the value of IT spillovers through *informal* channels. This may be due to data limitations; a common approach in this literature is to study the impact of spillover pools in which the weights through which the pools influence the focal entity are defined by geographical proximity, supply chain relationships, or competition (Alcácer and Chung 2007, Cheng and Nault 2007, Han et al. 2011, Cheng and Nault 2012). Whereas such approaches have their merits, they are unable to discern the specific channels of knowledge transfer and the nature of knowledge flows through firms.

As the context of our work differs from studies that examine the characteristics of knowledge in the R&D literature, it is unclear ex ante whether the results from that literature hold in our setting. The motivations behind the R&D literature focus on why and how prior R&D investments might help firms keep abreast of related technological developments and facilitate the assimilation of technology developed elsewhere (Tilton 1971, Cohen and Levinthal 1989). In contrast, in our setting—and in other prior research that studies knowledge transfer related to enterprise software—there is direct communication between a source and a recipient that should facilitate transfer of all types of knowledge (Szulanski 1996). As a result, prior information systems (IS) research argues that knowledge characteristics are less important than other antecedents to the transfer of knowledge in the context of enterprise information systems (Ko et al. 2005).

More broadly, our research approach provides a unique strategy to quantify heterogeneity in knowledge flows. In the past, because of the inherent measurement difficulties, efforts to measure this type of spillover rely upon survey-based measures (e.g., Bresnahan et al. 2002, Sambamurthy et al. 2003, Aral and Weill 2007), investments in related technologies (Greenstein and Nagle 2014, Nagle 2019), or more recently human capital data obtained from résumés and social network profiles (e.g., Tambe and Hitt 2014a). Similarly, most prior IS literature measures ACAP directly using survey data (Roberts et al. 2012). Challenges faced by many of these approaches is that they are often costly to implement, suffer from recall biases, and are limited in their ability to measure heterogeneity in knowledge flows. The research approach presented in this work provides insights for other researchers on how to use alternative, archival data to study questions in this research area.

2. Theory and Hypotheses

2.1. External Knowledge Flows, Absorptive Capacity, and Productivity

The effective implementation and use of IT within organizations emphasizes the view of IT as an enabler of business process innovation. Business process innovation requires a range of investments in computing hardware and software as well as changes to process flows, human capital, and other organizational practices (Bresnahan and Greenstein 1996). In the context of enterprise software, for example, adopters of enterprise resource planning (ERP) systems must incorporate local business rules into ERP software through a process of configuration and customization.

The knowledge and expertise of using IT to enable business process innovation is typically embodied in IT workers (e.g., Tambe and Hitt 2014a). This expertise can be accumulated through a process of on-the-job skill acquisition (Lieberman 1984, Benkard 2000, Thornton and Thompson 2001) or facilitated by accessing external knowledge sources. One channel through which knowledge can be transferred between firms is the direct acquisition of human capital through employment contracts. For example, through the acquisition of experienced IT workers, firms can obtain access to knowledge gained by these workers through their training at their previous employer (Tambe and Hitt 2014b).

This type of knowledge can also be transferred through informal interactions between firms, which are often labeled as knowledge spillovers (Griliches 1979). In the context of enterprise software, these types of informal interactions take place through many channels. For example, the Americas' SAP Users' Group hosts face-to-face meetings at which users can share experiences of implementing SAP software and benchmarking best practices. SAP also provides opportunities for knowledge transfer and human capital development via online channels. As described in greater detail as follows, SCN offers a platform for SAP users, partners, and employees to provide user-to-user support using web-based collaboration tools.

Recent evidence from other avenues for knowledge exchange, such as developer conferences, open source software development, and standards-setting processes, suggests that participation in such environments, whether virtual or physical, can augment the human capital of participants (Lakhani and von Hippel 2003, Nagle 2018, Foerderer 2020). In the context of developing and implementing IT systems, this type of human capital augmentation makes IT workers more productive when engaging in business process innovation, which, in turn, has a positive impact on firm productivity (Tambe and Hitt 2014a). In line with this thinking, we investigate whether exposure to external knowledge inputs through informal channels, such as online knowledge communities, has an impact on firm total factor productivity (TFP).

Firms may differ in their ability to assess the value of external IT knowledge and apply it for productive use. It is well understood in the context of R&D that outside sources of knowledge are an important input into the innovation process (Cohen and Levinthal 1990), and a firm's investment in R&D serves dual purposes: it not only generates new information, but also enhances the firm's ACAP (Cohen and Levinthal 1989). ACAP is shown to be path-dependent and is a function of prior knowledge accumulation (Cohen and Levinthal 1990). IS researchers broadly leverage the ACAP concept in the context of several streams of IS research, including research on knowledge transfer, IT assimilation, and IT business value (Roberts et al. 2012). Whereas many researchers identify ACAP as a capability, most prior work in IS argues that such capability is a function of relevant prior knowledge. In the context of IT-related spillovers, for example, ACAP is considered mainly dependent on a firm's prior IT investments (Han et al. 2011, Chang and Gurbaxani 2012a).

Extending this prior work, we first argue that IT investments serve dual purposes: IT not only creates value through the use of information systems, but also equips the firm with the ability to absorb external IT knowledge, assimilate it, and apply it for productive use, which contributes to the firm's productivity indirectly when opportunities to absorb external knowledge inputs arise. However, as is more fully articulated in the following section, we push this view further by capturing a neglected driver of ACAP in the existent IS literature: the nature of knowledge (Teece 1977, Kogut and Zander 1992) and its interplay with a firm's accumulated IT knowledge stock.

2.2. A Model of the Effect of IT Spillovers on a Firm's Productivity

To understand how IT spillovers can affect firm productivity, we adopt a production function approach and extend it by incorporating the effect of the knowledge stock related to enterprise systems as an input. A typical production function relates firm output to factors of input. For example, a simple form of a three-factor Cobb–Douglas production function is widely used in prior studies on IT productivity (Brynjolfsson and Hitt 1996, Dewan and Min 1997):

$$Y = AK^{\alpha}L^{\beta}C^{\eta}, \tag{1}$$

where Y is the quantity of production output, K is the stock of non-IT capital, L is the stock of labor, C is the stock of IT capital, and A denotes the TFP, which is defined as the output contribution that is not explained by the factor inputs and is often interpreted as technological progress. To incorporate the role of IT spillovers, we follow the literature on R&D spillovers by adding to Equation (1) a factor that captures the knowledge stock related to enterprise software, Z_{it}. In keeping with the work of Cohen and Levinthal (1989; henceforth C&L), we model the IT knowledge stock as

$$Z_{it} = M_{it}S_{it}^{\gamma_{it}}, \tag{2}$$

where M_{it} represents a firm's accumulated investments in enterprise systems, which includes investments in a combination of hardware, software, and human capital, such as training. The variable S_{it} represents flows of external knowledge available in the public domain that can be accessed through various channels as discussed.[7] The variable γ_{it} measures the extent to which the focal firm is able to recognize the value of external information, assimilate it, and effectively utilize it in a business setting. It, therefore, represents the firm's ACAP. We also assume a translog specification for ACAP, such that

$$\gamma_{it} = f(M_{it}, D_{it}) = \gamma_0 + \gamma_1 lnM_{it} + \gamma_2 lnD_{it} + \gamma_3 lnM_{it}lnD_{it}. \tag{3}$$

Consistent with C&L, Equation (3) specifies ACAP as a function of the firm's accumulated investments in enterprise systems, M_{it}; the characteristics of outside knowledge that make learning more difficult (or difficulty of learning), D_{it}; and their interaction. The interaction term reflects the idea that the importance of prior IT investments to knowledge assimilation depend upon the difficulty of learning, a key feature of the model in C&L.[8] We emphasize here that, whereas the dependence of γ_{it} on M_{it} is previously examined in the IS literature, the role of D_{it} and its interplay with M_{it}, to be explained as follows, are mostly ignored.

Integrating (2) and (3) into (1) and assuming the output elasticity of Z_{it} is φ, we can write a firm's output as

$$Y_{it} = K_{it}^{\alpha}L_{it}^{\beta}C_{it}^{\eta}M_{it}{}^{\varphi}S_{it}^{\varphi(\gamma_0+\gamma_1 lnM_{it}+\gamma_2 lnD_{it}+\gamma_3 lnM_{it}lnD_{it})}, \tag{4}$$

or in log form,

$$y_{it} = a + \alpha k_{it} + \beta l_{it} + \eta c_{it} + \varphi m_{it} + \gamma_0{}' s_{it} + \gamma_1{}' m_{it}s_{it} + \gamma_2{}' d_{it}s_{it} + \gamma_3{}' m_{it}d_{it}s_{it}, \tag{5}$$

where we use standard notation and denote terms in their log form using lowercase letters. Further, $\gamma_0' = \varphi\gamma_0$, and we use similar notation for the other knowledge absorption terms. We now use the log form of the production function (5) to derive a set of hypotheses related to how ACAP mediates the effect of IT spillovers on a firm's productivity.

2.2.1. Prior Related IT Investments.

Firms engaging in new business process innovation experience greater benefits when they have already made inroads through internal knowledge accumulation (Cohen and Levinthal 1989, Ko et al. 2005). The impact of prior knowledge in facilitating the absorption and adaptation of external knowledge to the focal firm's idiosyncratic needs is salient when some portion of that prior knowledge stock is related to that acquired externally (Cohen and Levinthal 1990). As pointed out above, this idea is consistent with a substantial body of work in IS (Roberts et al. 2012, Gao et al. 2017). We investigate the salience of this hypothesis by examining the extent to which IT expertise accumulated within an organization through prior investments in enterprise systems increases the productivity benefits of external knowledge acquisition.

More formally, our theory suggests that ACAP is increasing in a firm's accumulated investments in enterprise systems. That is, in Equation (3), $\partial\gamma_{it}/\partial m_{it} = \gamma_1 + \gamma_3 d_{it} > 0$. Because $\varphi > 0$ in (5), the positive dependence of ACAP on m_{it} implies that the effect of spillovers on output is positively moderated by m_{it}, or $\partial^2 y_{it}/(\partial s_{it}\partial m_{it}) = \gamma_1{}' + \gamma_3{}' d_{it} > 0$. We, therefore, formulate the following hypothesis:

Hypothesis 1. *The effect of IT spillovers related to enterprise software on a firm's output is positively moderated by the firm's prior investments in enterprise software.*

2.2.2. Characteristics of External Knowledge.

The costs of transferring knowledge across firm boundaries often depend upon the nature of knowledge (Teece 1977). C&L highlight that the value of external knowledge to a firm is greater when that knowledge is easier to assimilate and exploit. Whereas the idea is intuitive, specifying ex ante the features of external knowledge that affect learning in different contexts is more difficult and often less intuitive. C&L maintain that the assimilation of R&D knowledge depends on such factors as the complexity of the knowledge, how explicit and codified the relevant knowledge is, and the degree to which that outside knowledge is targeted to the needs of the firm (Cohen and Levinthal 1989, 1990). Their setting focuses on the ability of firms to incorporate findings from external R&D into internal R&D efforts; in contrast, ours involves adapting external knowledge on enterprise systems to a firm's specific needs. We highlight key institutional features of our setting to show how features of enterprise software map to the characteristics of difficulty as highlighted by C&L.

Earlier research shows that there are often significant knowledge barriers that firms must overcome to adopt and implement information systems such as enterprise software (Attewell 1992, McAfee 2002). The complexity of enterprise systems software and its high costs of deployment are well established (e.g., Davenport 2000) as is the need for exchange of tacit knowledge for its successful implementation (Ko et al. 2005). Here, we focus upon a particular attribute of the difficulty of knowledge that is related to the knowledge characteristics highlighted by C&L and that varies over time and across firms within our sample: the extent to which external knowledge is related to novel, newly developed technologies.

When knowledge is new, less information may be available on how to apply it properly (Cohen and Levinthal 1989, von Hippel 1994), and there may be causal ambiguity regarding why and when it provides a solution to problems (Szulanski 1996). Within the context of enterprise software, knowledge barriers to the implementation of enterprise software may arise both because of changes in existing products as well as when adopting new products. Existing products may change because of new version releases, mergers and acquisitions, technological progresses, and policy and regulatory changes (Foerderer et al. 2018). New products may incorporate novel knowledge domains or recent technological breakthroughs. As a result, much of the relevant knowledge for new products have yet to be codified. Further, standardization of the language used to describe new knowledge and the models used to represent it—an important aspect of knowledge codification—often takes time to mature (Cowan et al. 2000). All of these factors make new products difficult to learn.[9] This is consistent with studies showing that, for IT knowledge related to new applications, the extent of required coinvention may be greater and more context-dependent because standardized solutions have yet to be deployed and refined (Bresnahan and Greenstein 1996). Evidence of these differences is found in other settings as well; for example, von Hippel and Tyre (1995) show that avoidance of problems when using a new process machine may require a great deal of information about the setting in which it is to be applied.

Because of the significant knowledge barriers involved, we expect that, all else being equal, a firm's ACAP is lower when external knowledge is related to new or less mature technological domains. That is, in Equation (3), $\partial\gamma_{it}/\partial d_{it} = \gamma_2 + \gamma_3 m_{it} < 0$. Because $\varphi > 0$ in (5), the negative dependence of ACAP on d_{it} implies that the effect of spillovers on output is negatively moderated by d_{it}, or $\partial^2 y_{it}/(\partial s_{it}\partial d_{it}) = \gamma_2' + \gamma_3' m_{it} < 0$. We, therefore, formulate the following hypothesis:

Hypothesis 2. *The effect of IT spillovers on a firm's output is negatively moderated by features of external knowledge that make it more difficult to learn, such as its novelty.*

It should be noted that, whereas novel knowledge is more difficult to absorb, once absorbed, it can also have a greater effect on productivity. For example, during our sample period, knowledge of how to implement business intelligence and analytics software packages was more difficult to absorb because best practices had yet to be established and knowledge related to systems were unevenly distributed (Tambe 2014). However, acquisition of knowledge on how to use these systems was also likely to have a greater impact on productivity than that related to how to implement ERP packages, products that had already seen widespread adoption and for which best practices were widely established. As clarified in further detail subsequently, the net effect of these opposing forces depends upon prior investments in enterprise software.

2.2.3. The Interaction of Prior Related IT Investments and Characteristics of External Knowledge. The stock of related IT investments and the nature of knowledge acquired externally are characterized by important interdependencies. In particular, the role of prior IT investments in the process of knowledge assimilation depends upon the nature of knowledge the firm is seeking to acquire. A related point has been demonstrated in the R&D literature (Cohen and Levinthal 1989), which shows that related internal R&D becomes more important in the acquisition of external knowledge when that external knowledge is more complex and less targeted at the needs of the firm. Under these circumstances, the firm's prior related investment becomes vital to the assimilation and use of external knowledge.

As discussed, some characteristics make external knowledge more difficult to learn, which, in turn, might influence the extent to which prior related IT investments contribute to a firm's absorptive capacity. We expect that prior related IT investments play a greater role in absorbing knowledge acquired from external sources when a large fraction of external knowledge involves novel knowledge. In other words, firms with prior related IT investments receive greater benefits (i.e., higher productivity) from external knowledge flows related to novel knowledge—knowledge that is difficult to be transferred and absorbed. This is because it is more challenging for the receiving firm to translate insights gained from this type of knowledge acquisition into a valuable set of actions related to processes, decision rights, and organization. In other words, firms with prior related investments on how to implement enterprise software are able to derive value from inflows from this type of knowledge—in this setting, firms are able to put insights learned into productive use. However, prior IT investments and associated cumulated IT human capital have less influence on the value obtained from flows of less novel and well-codified knowledge because transferring such knowledge requires little adaptation and customization.

To summarize, we expect prior investments in enterprise systems to be more important to the firms' ability to exploit external knowledge when the external knowledge is novel. Referencing Equation (3), we expect that $\partial^2 \gamma_{it}/(\partial m_{it} \partial d_{it}) = \gamma_3 > 0$. Because $\varphi > 0$ in (5), the positive interaction between m_{it} and d_{it} on ACAP implies that $\partial^3 y_{it}/(\partial s_{it} \partial m_{it} \partial d_{it}) = \gamma_3' > 0$. We, therefore, formulate the following hypothesis:

Hypothesis 3. *When the difficulty of learning is high, such as when external knowledge is novel, the moderating effects of prior investments in enterprise software on the relationship between IT spillovers and a firm's output are stronger.*

3. Research Context

Our research questions require a robust measure of interfirm knowledge flows related to the use of IT with observable knowledge characteristics. We use the online community network created by SAP as the context of our study.

We choose enterprise software as the background for measuring IT knowledge flows for several reasons. First, investment in enterprise software and its implementation accounts for a significant portion of total business-related IT spending (Brynjolfsson et al. 2002). According to one estimate, in 2013, SAP customers across the world invested around $204 billion dollars in SAP-related software, labor, and infrastructure (e.g., Mirchandani 2014). In addition, adoption of enterprise software is shown to be associated with significant improvements in firm financial and operational performance (Hitt et al. 2002). However, implementing enterprise software is complex and requires complementary business process innovation; because of these challenges, projects frequently take longer than expected, and benefits take a long time to achieve (McAfee 2002). Finally, knowledge of how to implement enterprise software systems is unevenly distributed among users (Yusuf et al. 2004) and, because of the heterogeneous environments in which systems are implemented, not easily contracted out. Internal human capital accumulation occurs as users learn how to deploy software functionality in their organizations through a series of projects (Walker 2012). Thus, our environment offers a useful test case for understanding the interrelationships between the internal stock of IT knowledge and efforts to develop human capital through external interactions facilitating the accumulation of internal knowledge.

In 2003, SAP established an internet-based network of practice, the SAP Developer Network (SDN). The SDN was later expanded to include a community for business process experts and was expanded still further over time to incorporate other communities that interface with SAP's products. Given this increase in breadth, the SDN is now known as SCN. It hosts forums, expert blogs, a technical library, article downloads, a code-sharing gallery, e-learning catalogs, wikis, and other facilities through which users contribute their knowledge. As of 2008, the community comprised active users from 224 different countries.

The SCN community has a contributor recognition program that awards points to community users for contributing technical articles, code samples, videos, wiki entries, forum posts, and weblogs. For example, when users participate in a forum discussion, they can receive points for posting solutions to existing discussion threads marked as questions if their answers are deemed helpful by the person who asks the question. SAP publicly recognizes its most active contributors. For example, on the "Top Contributors" page, SCN lists the top 50 contributors as measured by total reward points.

Participation in the community network starts with a registration process in which a user builds a profile by providing basic personal information, such as the name of the user's employer. Using this piece of information, it is possible to aggregate the knowledge flows to firms whose employees actively participate in SCN. The user's profile also lists the user's name, country of origin, relationship to SAP, email address, phone number, expertise, and LinkedIn profile page.

To track knowledge flows among SCN users, we focus on user interactions through the most frequently used communication format: discussion forums. Although SCN users may access knowledge through other formats, such as wikis, blogs, and articles, these other formats have fewer active participants than discussion forums, and knowledge flows arising from the use of these other channels are unfortunately not measurable. The primary purpose of the discussion forums is to provide an avenue for conversations among the community users to help one another solve problems encountered during the implementation, deployment, and use of SAP software (Fahey et al. 2007). The forums are organized according to domains of knowledge or expertise, each of which usually corresponds to a technical domain (e.g., database or operating system), a particular SAP software module, or the application of SAP to a particular industry.

Conversations in each forum are organized by discussion *threads*. Each thread is initiated by a knowledge seeker, who posts a specific question in a topic forum. Knowledge contributors, in turn, post responses that attempt to answer the question. A discussion thread, therefore, consists of a list of *messages*, and each message (either a question or an attempted answer) contains the information about the member who posts the message, the body of the message, and a time stamp. After a correct answer (judged by the knowledge seeker) is received, the discussion thread is closed.

We developed a web scripting tool and obtained the complete history of SCN forum discussions from 2004 to 2008. The data set includes about 1.1 million discussion threads with 5.0 million messages posted in 209 topic forums. In Appendix Table A.1, we present some summary statistics of the evolution of the SCN over our sample period, including numbers of registered users, topic forums, and the discussion threads posted in these forums. Overall, we find that the online community has experienced rapid growth since its establishment: by the end of our sample, roughly one quarter of the questions are solved by the collective effort of the community users, and the average time to obtain a correct solution is less than five days.

4. Data and Methods

4.1. Estimation Model

We estimate Equation (5) after introducing firm and year fixed effects and the idiosyncratic error, using a panel data model exploiting within-firm variation over time as specified in the following equation:

$$y_{it} = a + \alpha k_{it} + \beta l_{it} + \eta c_{it} + \varphi m_{it} + \gamma_0{}' s_{it} + \gamma_1{}' m_{it} s_{it} + \gamma_2{}' d_{it} s_{it} + \gamma_3{}' m_{it} d_{it} s_{it} + \mu_i + \delta_t + \varepsilon_{it} \quad (6)$$

As explained in Section 2.2, Hypothesis 1 is equivalent to $\partial^2 y_{it}/(\partial s_{it} \partial m_{it}) = \gamma_1' + \gamma_3' d_{it} > 0$. A test of Hypothesis 2 is equivalent to testing $\partial^2 y_{it}/(\partial s_{it} \partial d_{it}) = \gamma_2' + \gamma_3' m_{it} < 0$. Finally, Hypothesis 3 is equivalent to testing $\gamma_3' > 0$.

4.2. Data

We construct a data set of publicly traded firms that are SAP adopters. Our data come from a variety of sources. Our primary measure of knowledge transfer comes from user activities in the discussion forums on the SCN. To identify SAP adopters, we obtained a detailed list of all installations of SAP product modules in the United States prior to the end of 2004 from SAP. We use the Harte Hanks Computer Intelligence (CI) Technology database to collect firm-level IT investment data. The CI database records detailed information about IT infrastructure for most of the Fortune 1,000 firms, including data on the quantity of mainframes, peripherals, minicomputers, servers, and PC systems as well as other IT hardware. The CI database is widely used by prior studies to investigate issues related to IT productivity (e.g., Brynjolfsson and Hitt 2003, Dewan et al. 2007, Chwelos et al. 2010). The CI data were matched with Standard and Poor's Compustat database to obtain financial data that we use to construct measures of production output, non-IT capital stock, and labor expenses.

4.2.1. Sample. Our sample is constructed in several steps. It begins in 2004 with the start of SCN and ends in 2008, which is the last year for which we have IT investment data. To obtain the data for our sample, we first retrieve the set of firms that were on the Fortune 1,000 list at least once during 2004–2008 and match them to Compustat data. We then match these firms with the CI database. Because we are interested in knowledge spillovers related to the implementation and the related business process innovation in SAP products, we further restrict our sample to those firms that had installed at least one SAP module prior to the end of 2004. We note that this sample of firms represents the complete set of firms for whom our hypotheses are relevant. Our interest is on how the effect of knowledge flows related to SAP software on productivity is affected by the difficulty of knowledge (d_{it}) and prior related investments (m_{it}), and these key variables in our regression model such as d_{it} and m_{it} are not defined for firms without knowledge flows related to SAP or investments made in SAP software. The final sample is an unbalanced panel of 275 firms with 1,240 observations over a five-year period.

4.2.2. Variables. In this section we describe the variables used in our analysis. We first describe the variables that measure productivity, labor, and IT and non-IT capital, followed by the variables measuring knowledge flows and those related to absorptive capacity.

4.2.2.1. Production Function Inputs and Outputs. *IT capital*: Our measure of IT capital is derived from the CI database. The information in the database covers major categories of IT hardware investments made by firms, such as personal computing, systems and servers, networking, software, storage, and managed services. Historically, the CI database has provided direct measures of IT capital stock, but this measure is not available over the years of our sample. As a proxy, we adopt the method used by Brynjolfsson and Hitt (1995), Dewan and Min (1997), Gu et al. (2008), and Hitt and Brynjolfsson (1996) and measure the IT capital stock using an estimate of the market value of the IT hardware systems plus three times the current year's IT labor expenses. Inclusion of IT labor expense in the calculation of IT capital is justified by the fact that a large fraction of IT labor expenses is dedicated to the development of computer software, which is a capital good. The assumption that underlies this method is that the current IT labor spending is a good proxy for the IT labor expenses in the recent past, and IT staff "stock" depreciates fully in three years (Brynjolfsson and Hitt 1995). Details of the computation of this variable are presented in Appendix 1.

Production output: We follow prior literature (Dewan and Min 1997, Brynjolfsson and Hitt 2003) and use added value as the measure of production output, which equals deflated sales less deflated materials. Compared with sales, added value is said to be less noisy and more comparable across industry sectors (Dewan and Min 1997). Annual sales numbers are retrieved from Compustat, and we deflate them using industry-specific (at the two-digit North American Industry Classification System (NAICS) sector) price deflators from the Bureau of Economic Analysis' Gross Output and Related Series by Industry. Materials are calculated by subtracting undeflated labor and related expenses (Compustat data item XLR) from undeflated total operating expenses (Compustat data item XOPR), and deflating by the Bureau of Labor Statistics (BLS) Producer Price Index for intermediate materials, supplies, and components.

Non-IT capital: The calculation of total capital stock is similar to that in Brynjolfsson and Hitt (2003) for ordinary capital. Specifically, the gross book value of capital stock (property, plant, and equipment (total gross), Compustat data item PPEGT) is deflated by an industry-specific capital investment deflator reported in BLS 1987–2010 Detailed Capital Measures.[10] In order to apply the deflators, the average age of capital stock is calculated as the ratio of total accumulated depreciation (Compustat data item DPACT) to current depreciation. We then subtract the deflated computer capital from deflated total capital to get the value of non-IT capital.

Non-IT labor: Consistent with prior studies on IT productivity (Bresnahan et al. 2002, Brynjolfsson and Hitt 2003), total labor expense is either obtained directly from Compustat labor and related expenses (data item XLR) or calculated as the product of a firm's reported number of employees (Compustat data item EMP) and industry-average labor cost per employee and deflated by the BLS Employment Cost Index (ECI) for private industry workers. Average labor cost per employee is obtained from national sector NAICS industry-specific estimates series of BLS occupational employment and wage statistics (OES). To account for the fraction of benefits in total compensation, we multiply the wage number by the ratio of total compensation to salary, which is obtained from BLS Employer Costs for Employee Compensation (ECEC) series. Non-IT labor is defined as the difference between deflated total labor expense and IT labor expense.

4.2.2.2. Variables Measuring Knowledge Flows and Absorptive Capacity. We measure flows of IT-related knowledge acquired externally, S_{it}, from forum conversations that took place on the SCN. For each question that is posted, the rules of the SAP reward program specify that the knowledge seeker can use discretion to judge the quality of answers posted by knowledge contributors and distribute reward points as follows: 10 reward points for correct answers (at most one answer can be evaluated as correct), 6 points for very helpful answers (at most two answers), and 2 points for helpful answers (no limit on number). We define a knowledge inflow as an incident when a knowledge seeker gives reward points to knowledge contributors in recognition of their quality responses. As noted, we use a crawler program to identify user information, such as location and firm. Next, we select all the users that reside in the United States and match them to firms in our sample by examining their employer affiliations and domains of their email addresses.

For each user a who is an employee of firm i, we retrieve all the discussion threads that were initiated by a in year t and examine the history of the answers posted by other forum users. If a received any correct, very helpful, or helpful answers in year t, the total number of reward points a gave to the knowledge contributors are used as a proxy for inward IT spillovers to a. The reward points are then aggregated across all the threads posted by a in year t to derive an individual-level knowledge inflow, S_{at}. The firm-level spillover variable is defined as the sum of knowledge inflows of all the individuals who are employees of the firm:

$$S_{it} = \sum_{a \in F_i} S_{at},$$

where F_i is the set of users who are employees of firm i.[11] We exclude from this measure within-firm knowledge flows, that is, knowledge flows in which both the source and the recipient are employed by the focal firm i.

Our measure of knowledge inflow is likely to suffer from measurement error because of missing data on the knowledge seekers who did not report their employers.[12] However, we observe no systematic differences in knowledge inflow between questions asked by knowledge seekers who reported their employers and those asked by seekers who did not reveal their employers: the average inflow per question per year is 2.91 for nonreporting seekers and 3.08 for reporting seekers, and the difference is not statistically significant (p = 0.35). If firms strategically promote employee activity in SCN and other communities (Mehra et al. 2011), then S_{it} may serve as a proxy for the broader receptivity of the firm to external inflows. We consider this possibility further in Section 5.2.

Our primary measure of difficulty of learning, D_{it}, captures the novelty of external knowledge and the rate at which it is changing. As noted earlier, novelty may arise from two sources. One source of novelty is changes in existing products that may arise from new versions, major upgrades or changes to the underlying technology used in the product. For example, during our sample period, SAP switched from its traditional product strategy to a platform ecosystem strategy when it unveiled its NetWeaver platform, which incorporated its traditional proprietary ERP technologies with more recent web-based technologies (Lakhani et al. 2014). Another source may arise from the introduction of entirely new products, such as, the introduction of new products and services around Business Objects after SAP's acquisition of that company. New forums are introduced as a result of these developments. For example, after SAP's acquisition of Business Objects, it merged the Business Objects Diamond community into SCN[13] and introduced several new forums, such as SAP BusinessObjects Enterprise/Edge and SAP Crystal Reports Server Administration. Following the launch of the NetWeaver-based platform, a total of 18 new forums related to NetWeaver were created.

We construct D_{it} as the percentage of SCN knowledge flows that are derived from forums that are less than one year old at the time the flow takes place. The variable, D_{it}, is defined as (1 + knowledge flows from new forums)/(1 + total knowledge flows), where one is added to both the numerator and denominator to avoid taking the log of zero when the variable d_{it}—the log of D_{it}—is entered into the regression. Summary statistics suggest that questions posted in new forums receive fewer replies and are less likely to be solved. For example, the number of replies posted within three days is equal to 2.660 for new forums compared with 3.777 for existing forums ($p < 0.001$), whereas the likelihood of receiving an answer within 10 days that the question asker says has solved the problem is equal to 16.7% for new forums and 23.9% for existing forums ($p < 0.001$). Whereas we are unable to discern whether these differences are due to something inherent about the nature of knowledge in new and existing forums or because the number of participants in new forums is smaller, in either case, it means that, other things being equal, questions posted in new forums are less likely to receive answers that address the firm's needs.

We explore the robustness of our analysis to another measure that captures the degree to which knowledge is difficult to assimilate and exploit. Prior research emphasizes two distinct dimensions of IT knowledge that are particularly relevant in the process of adopting an information system: technical knowledge and business functional knowledge (Lee et al. 1995). The latter type of knowledge is often context-dependent and requires identifying the correct system of activities within the context of the firm and implementing them successfully (Brynjolfsson and Milgrom 2012). We distinguish between these two sources of knowledge by examining the forum in which a knowledge seeker's question is raised. We define a forum as technically oriented if the forum is dedicated to topics related to low-level, enabling technologies of an enterprise system, such as programming languages, database technologies, data transfer issues, and reporting and formatting tools. In contrast, we define a forum as business-oriented if the discussion topics in the forum focus on the configuration of the enterprise system to implement a particular business function or process, such as monitoring employee performance, coordination of supply chains, consolidating procurement processes, or managing projects.[14] Our alternative measure for D_{it} is, therefore, defined as the share of the inflows related to business functional knowledge.

Unlike our variables S_{it} and D_{it} for which we have direct measures of knowledge flows and their characteristics, we do not directly observe prior investments in enterprise systems, M_{it}. As is well known, measures of direct (e.g., software license fees) and indirect (e.g., human capital investments) spending related to enterprise systems generally cannot be observed except through survey measures, such as those employed by Brynjolfsson et al. (2005). In the absence of direct measures of M_{it}, we compute proxies based on data obtained from inside and outside the SCN forums.

Our primary measure of M_{it} incorporates two important elements of investments in enterprise software: the extent of enterprise software adoption (which is directly related to software licensing costs and implementation costs), and human capital investments related to its adoption, such as training costs. A typical SAP system consists of a series of technical and functional modules.[15] Using data that we obtained from SAP, we measure enterprise software adoption by calculating the number of SAP modules that were installed by the focal firm prior to 2004 (the first year of our sample). We then use data from the CI database on the number of IT employees in a firm as a proxy for human capital investments. We assume that $M_{it} = (IT\ employee)^{(\alpha \times number\ of\ SAP\ modules)}$ or, equivalently, $m_{it} = log(M_{it}) = \alpha \times (number\ of\ SAP\ modules) \times log(IT\ employee)$.

We additionally create an alternative measure of prior related knowledge based on the participation of the firm's employees in the SCN community. Specifically, we compute the *cumulative* contributions to SCN forums made by all the employees of firm i prior to year t (measured by reward points they earned) and create a binary variable whose value is set to one if the cumulative contribution made by a firm's employees is greater than the sample mean in year t. These types of contributions to crowdsourced communities are shown to contribute to organizational learning (Nagle 2018).

In sum, we view these measures as related and use them together to triangulate our understanding of the behavior of the same (ultimately unobserved) variable. An analysis of the data supports this assertion: the mean of the first measure of m_{it}—the IT employee weighted number of SAP modules—for firms that have a high cumulative knowledge contribution is 43% higher than that for firms with a low cumulative knowledge contribution.

4.2.2.3. Control Variables. We also include a number of variables that control for firm activities on the SCN forum other than knowledge flows, such as the cumulative number of registered users who are the focal firm's employees in the SCN, and the total number of questions raised by a firm's employees. Although not directly related to receiving answers to questions, these variables could capture other unobservable factors associated with the propensity to learn or use the online platform, such as heterogeneity across firms' policies on the use of SCN or other relevant firm capabilities.

One possible source of omitted variable bias is our limited ability to observe other forms of knowledge inflows associated with user activities on SCN. In particular, we measure spillovers based on forum Q&A discussions, and for such spillovers to be observed, the knowledge seeker must explicitly ask a question in the forum and receive some helpful answers. However, knowledge seekers may also obtain knowledge spillovers without explicitly asking questions, especially when similar problems have already been solved by community members—for example, they can perform a keyword search on SCN forums and find existing solutions to their problems instead of initiating a new Q&A discussion thread. To address this confounding factor, we construct a control variable to capture the effect of *learning by reading* existing posts. The first step in this process involves identifying the size of the existing knowledge pool associated with a forum j in year t. We count the number of all resolved cases (questions that received correct answers) by the end of year t in forum j—defined as P_{jt}—as a proxy. To account for the degree to which firm i is able to use the existing knowledge pools (e.g., by reading existing posts), in the second step, we use the share of firm i's activity in forum j as weight. We experiment with two different weighting schemes: $w_{ijt} = (u_{ijt}/(\sum_j u_{ijt}))$, where u_{ijt} is the number of firm i's employees who were active in forum j and year t (a user is active if the user participated in at least one discussion in forum j and year t), and $w'_{ijt} = (q_{ijt}/(\sum_j q_{ijt}))$, where q_{ijt} is the number of questions raised by firm i's employees in forum j and year t. The control variable is then defined as $\sum_j P_{jt} \times w_{ijt}$. The two weighting schemes yield very similar results when the control variable is added to the regressions. For brevity, we report the result using the second weighting scheme.

Table 1 reports the summary statistics of the variables. The average firm in the sample has sales of \$16.65 billion, added value of \$5.49 billion, and 41,167 employees, consistent with our sample being large, publicly traded, Fortune 1,000 SAP adopters. In addition, firms in our sample invest heavily in IT capital, which has a mean level of \$97.49 million and a maximum of \$1.18 billion. Table 2 provides the correlation matrix among the key variables. In Table A.2, we provide a breakdown of the sample firms by vertical industries, which is based on two-digit NACIS sectors. It is notable that manufacturing firms account for the majority (66%) of the sample, followed by utilities (8%).

5. Results

5.1. Test of the Absorptive Capacity Model and Hypotheses

We present our main results in Table 3; all models include firm and year fixed effects. We note that, in Equation (6), the term d_{it} does not directly influence productivity, that is, its effect on productivity operates only through the influence of external knowledge flows, s_{it}. As noted, because it is calculated based upon the characteristics of external knowledge flows, d_{it} is defined only when $s_{it} > 0$. As a result, we include it in our regression models only as it appears in Equation (6), that is, in the interaction terms of $d_{it}s_{it}$ and $m_{it}d_{it}s_{it}$.

Before we show the result from the full ACAP model in column (5), column (1) presents a model in which we assume that firms have a homogeneous absorptive capacity, that is, γ_{it} is a constant independent of m_{it} and d_{it}. Because it is omitting key terms that influence the relationship between spillovers and productivity, it is subject to a specification error. However, we include it to construct a basis against which we can compare our primary results. The results from this model show a positive effect of knowledge flows, implying that a 1% increase in the amount of inward knowledge flows is associated with 0.01435% increase in the added value produced by a firm. Considering that the added value of an average firm in our sample is \$5.491 billion, this translates into a \$0.79 million increase in production output. To put it another way, for the average firm in our sample, doubling the amount of external knowledge obtained from SCN

Table 1. Summary Statistics

Variable	Mean	Standard deviation	Minimum	Maximum
Annual sales, million $	16,649.51	33,024.88	298.91	364,392.40
Added value, million $	5,491.17	8,811.34	118.11	73,242.29
Non-IT capital, million $	12,526.01	29,661.25	48.44	321,772.70
IT capital, million $	97.49	138.29	0.00	1,181.67
Non-IT labor, million $	2,781.93	4,405.87	28.75	40,586.13
Number of employees, thousands	41.17	59.59	0.66	428
Knowledge inflows, reward points	10.71	93.82	0	2,190
Difficulty of learning, % knowledge related to business[a]	0.43	0.40	0.00	1
Difficulty of learning, % knowledge from new forums[a]	0.11	0.21	0.00	1
Prior investments in enterprise systems in log	81.67	54.19	0	291.11
high related knowledge in human capital, binary	0.08	0.26	0	1
Number of SCN users	2.94	6.52	0	97
Number of questions	3.58	12.45	0	220
Learning by reading	11,723.21	93,621.67	0	2,451,835

Notes. Number of observations: 1,240. Number of firms: 275.
[a]Summary statistics for difficulty are based on observations with nonzero knowledge flows.

(i.e., knowledge inflows moving from the sample mean, 10.71 points, to 21.42 points) increases added value from $2.683 billion to $2.710 billion: a $27 million increase.

In the remaining columns, we explore the implications of incorporating the elements of ACAP. In keeping with recent work that has sought to understand how interdependencies between organizational characteristics, IT, and the external orientation of a firm can contribute to productivity (e.g., Aral et al. 2012, Tambe et al. 2012, Nagle 2018) we begin by examining the impact of changes of m_{it} and d_{it} separately (i.e., without accounting for their interdependence) before estimating the full ACAP model as specified in Equation (6). These results are included in columns (2)–(4), in which we incrementally include first the terms $m_{it} \times s_{it}$, $d_{it} \times s_{it}$, and then both terms together (but excluding the term $m_{it} \times d_{it} \times s_{it}$). In such models, when ignoring the higher order interaction terms, the impact of changes of the lower order terms are typically close to their marginal effects when evaluated at the mean value of the omitted covariates (Balli and Sørensen 2013). Presenting them provides a useful comparison against the full model. However, omitting these higher order terms, whose "true" effects are nonzero from the equation, biases the lower order coefficients (Aiken et al. 1991). Further, they do not allow us to measure the interdependencies among the ACAP terms, a key contribution of this paper.

We next present the results of the full ACAP model, accounting for the roles of m_{it} and d_{it} and their interdependencies, in column (5) of Table 3. In this model, a test of our hypotheses requires us to compute the linear combinations of coefficients as described in Section 2.2, and we present the test of hypotheses in Table 4. A test of Hypothesis 1 in these models represents a test of the moderating effect of prior investments (m_{it}) on the relationship between knowledge flows (s_{it}) and

Table 2. Pearson Correlation Matrix of the Variables

	Variable	1	2	3	4	5	6	7	8	9	10	11	12	13
1	*Annual sales*	1.0000												
2	*Added value*	0.8612*	1.0000											
3	*Non-IT capital*	0.8313*	0.7690*	1.0000										
4	*IT capital*	0.3398*	0.4847*	0.3327*	1.0000									
5	*Non-IT labor*	0.5479*	0.8101*	0.3659*	0.5767*	1.0000								
6	*Number of employees*	0.5236*	0.7570*	0.3716*	0.5557*	0.9335*	1.0000							
7	*Knowledge flows*	0.0166	0.0443	−0.0014	−0.0023	0.0593*	0.0349	1.0000						
8	*Difficulty–new forums*	−0.0867*	−0.1509*	−0.0387	−0.0457	−0.1664*	−0.1066*	−0.4058*	1.0000					
9	*Difficulty–business*	−0.1048*	−0.1612*	−0.0415	−0.0610*	−0.1799*	−0.1241*	−0.3749*	0.8042*	1.0000				
10	*Prior investments in ES*	0.0545	0.0262	0.0488	0.0005	0.0208	0.0152	0.2277*	−0.0483	−0.0668	1.0000			
11	*High human capital*	0.1052*	0.1454*	0.0121	0.0164	0.1479*	0.1076*	0.1235*	−0.3367*	−0.2799*	0.1561*	1.0000		
12	*Users*	0.2143*	0.3408*	0.1140*	0.1380*	0.3918*	0.3023*	0.1858*	−0.4509*	−0.2764*	0.3614*	0.4385*	1.0000	
13	*Questions*	0.1090*	0.2071*	0.0323	0.0596*	0.2626*	0.1868*	0.8117*	−0.6436*	−0.4651*	0.2373*	0.3570*	0.6123*	1.0000
14	*Learning by reading*	0.0193	0.0470	−0.0029	−0.0003	0.0679*	0.0419	0.9185*	−0.4008*	−0.2640*	0.0820*	0.1383*	0.2623*	0.7974*

*$p < 0.05$.

Table 3. Baseline Absorptive Capacity Models

Variables	(1) With homogenous ACAP	(2) With $s_{it} \times m_{it}$	(3) With $s_{it} \times d_{it}$	(4) With both $s_{it} \times m_{it}$ and $s_{it} \times d_{it}$	(5) Full ACAP model with three-way interaction	(6) Subsample: Firms with positive knowledge flows
k_{it}	0.11181**	0.11263**	0.11242**	0.11325**	0.11389**	0.04089
	(0.04415)	(0.04426)	(0.04416)	(0.04426)	(0.04427)	(0.09505)
c_{it}	0.01908***	0.01836**	0.01909***	0.01836**	0.01810**	0.02358*
	(0.00735)	(0.00736)	(0.00734)	(0.00736)	(0.00738)	(0.01216)
l_{it}	0.72779***	0.72767***	0.72762***	0.72750***	0.72692***	0.79400***
	(0.05871)	(0.05875)	(0.05875)	(0.05879)	(0.05877)	(0.12264)
m_{it}	−0.00065	−0.00064	−0.00065	−0.00065	−0.00064	−0.00068
	(0.00070)	(0.00070)	(0.00070)	(0.00070)	(0.00070)	(0.00056)
s_{it}	0.01435**	0.00649	0.00893	0.00099	−0.00838	−0.00398
	(0.00655)	(0.00720)	(0.00737)	(0.00845)	(0.01020)	(0.01155)
$s_{it} \times m_{it}$		0.00007*		0.00007*	0.00015***	0.00014**
		(0.00004)		(0.00004)	(0.00005)	(0.00007)
$s_{it} \times d_{it}$			−0.00153	−0.00155	−0.00404**	−0.00373*
			(0.00119)	(0.00122)	(0.00193)	(0.00203)
$s_{it} \times d_{it} \times m_{it}$					0.00002**	0.00002
					(0.00001)	(0.00001)
Log(registered users)	−0.01243	−0.01434	−0.01208	−0.01400	−0.01399	−0.03354
	(0.01575)	(0.01604)	(0.01566)	(0.01595)	(0.01592)	(0.02554)
Log(questions)	−0.01165	−0.01178	−0.01258	−0.01272	−0.01282	−0.03212
	(0.02065)	(0.02066)	(0.02065)	(0.02064)	(0.02061)	(0.02518)
Log(learning by reading)	0.00080	0.00104	0.00092	0.00116	0.00119	0.00619
	(0.00415)	(0.00417)	(0.00417)	(0.00418)	(0.00418)	(0.00600)
Constant	1.67671***	1.67324***	1.67284***	1.66934***	1.66744***	1.99911
	(0.45722)	(0.45771)	(0.45763)	(0.45810)	(0.45819)	(1.21803)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,240	1,240	1,240	1,240	1,240	275
R^2	0.57767	0.57814	0.57792	0.57840	0.57863	0.64478
Number of firms	275	275	275	275	275	58

Notes. Unless otherwise noted, k_{it} = log(non-IT capital), c_{it} = log(IT capital), l_{it} = log(non-IT labor), m_{it} = log(prior related investment), s_{it} = log(knowledge flows), d_{it} = log(difficulty of learning). The dependent variable is the natural logarithm of value added. All models use firm-level fixed effects and year dummies. Robust standard errors (clustered by firm) are in parentheses. All R^2 values are "within" estimates that do not include the explanatory power of the fixed effects.

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

output, holding difficulty of learning (d_{it}) at mean values. Thus, it is a test of the statistic that $(\gamma_1' + \gamma_3' d_{it}) > 0$.[16] Unless specified otherwise, we focus our discussion on the baseline estimates of column (5) of Table 4. The test in column (5) shows that Hypothesis 1 is supported at the $p < 0.01$ level. The point estimate of this hypothesis test is similar to those of the models excluding the higher order terms (0.00014 in column (5) of Table 4 compared with 0.00007 in columns (2) and (4)) though the magnitude and significance levels are slightly higher, perhaps because the fully specified model incorporates the impact of the (nonzero) higher order terms.

We further illustrate the moderating effect of prior investments on spillovers by plotting the value of the statistic $(\gamma_1' + \gamma_3' d_{it})$ at different levels of d_{it}, together with its 90% confidence interval, in Figure 1. The figure shows that prior investments increase the effect of spillovers on output for most of the mass of data in our sample and have statistically significant effects for values of d_{it} at or greater than the median.

We next examine the moderating effect of the characteristics of knowledge. Hypothesis 2, which states that the effects of inward knowledge flows on productivity are smaller when those knowledge flows are more difficult to learn, such as when external knowledge is novel, is supported at the $p < 0.1$ level in column (5) of Table 4. As is the case in our earlier discussion of Hypothesis 1, columns (3) and (4) of Table 4 show that the point estimates of the test statistic of Hypothesis 2 in these models are similar to those of our baseline (fully specified, column (5)) model but with slightly lower economic and statistical significance.

We further illustrate the value of the statistic $(\gamma_2' + \gamma_3' m_{it})$ at different levels of m_{it}, together with its 90% confidence interval, in Figure 2. These results

Table 4. Hypotheses Testing

Hypothesis	(1)	(2)	(3)	(4)	(5)	(6)
Hypothesis 1	N/A	0.00007 (p = 0.094)	N/A	0.00007 (p = 0.099)	0.00014 (p = 0.007)	0.00016 (p = 0.053)
Hypothesis 2	N/A	N/A	−0.00153 (p = 0.197)	−0.00155 (p = 0.205)	−0.00233 (p = 0.075)	−0.00171 (p = 0.158)
Hypothesis 3	N/A	N/A	N/A	N/A	0.00002 (p = 0.031)	0.00002 (p = 0.103)

Notes. In columns (5) and (6) Hypothesis 1 is tested by computing $(\gamma_1{}' + \gamma_3{}' d_{it}) > 0$ with d_{it} at mean. In columns (5) and (6) Hypothesis 2 is tested by computing $(\gamma_2' + \gamma_3{}' m_{it}) < 0)$ with m_{it} at mean.

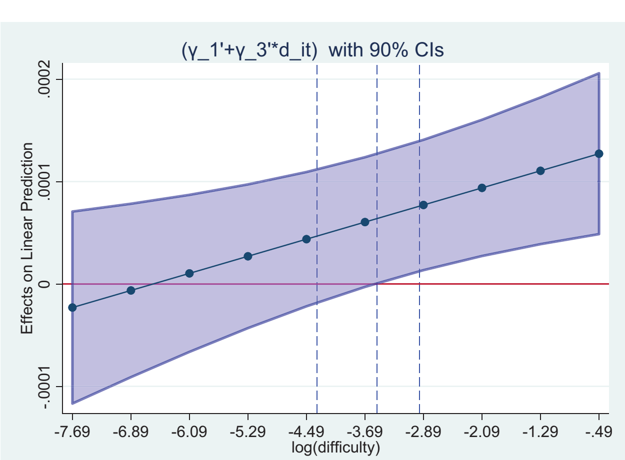
highlight the critical interdependence between d_{it} and m_{it}. Whereas the point estimate of the hypothesis test of the marginal effects of d_{it} were supported at only at the $p < 0.1$ level when evaluated at mean values of m_{it}, d_{it} has a statistically significant and negative impact on the value of spillovers when m_{it} is at or below the sample median. For example, when m_{it} is at the 25th percentile, the value of the statistic $(\gamma_2' + \gamma_3{}' m_{it})$ is −0.00316 ($p < 0.05$). When prior investments are sufficiently high, however, increases in d_{it} do not impede spillover benefits.

Finally, we assess the interaction effect of prior investments in enterprise software and difficulty of learning on the returns of knowledge inflows. The test of Hypothesis 3 can be performed directly by examining $\gamma_3' > 0$ in regression Equation (6). We observe a positive and significant coefficient estimate of the three-way interaction $s_{it} \times d_{it} \times m_{it}$ at the $p < 0.05$ level, supporting Hypothesis 3. That is, prior IT investments play a greater role when external knowledge is difficult to learn.

Consistent with the theory of absorptive capacity, we observe that prior related investment in enterprise software serves dual purposes: beyond its direct contribution to productivity, it also contributes to productivity indirectly by enhancing a firm's IT-related absorptive capacity, thereby allowing the firm to identify and exploit external knowledge. To quantify the economic implications of absorptive capacity for the value of knowledge inflows, we compute the output elasticities of knowledge flows for an average firm in the sample. Because our measure of difficulty of learning (d_{it}) is only defined for firms for which $S_{it} > 0$, we compute the effects of knowledge flows evaluated based upon the mean values of m_{it} and d_{it} conditional on $S_{it} > 0$. Evaluating the marginal effect at this point and based on the regression results in column (5), we find that a 1% increase in the amount of inward knowledge flow is associated with a 0.0042% increase in the added value produced by the firm ($p < 0.05$), which translates to a \$0.23 million increase in added value. Of course, the effects of knowledge inflow are even greater when m_{it} (prior investments) is higher and/or d_{it} (difficulty of learning) is lower. For example, marginal effect calculations suggest that when m_{it} is at the third quartile of the sample (and d_{it} is at its mean value), the output elasticity of knowledge flows is 0.01437 ($p < 0.01$), and a 1% increase of knowledge flow leads to a \$0.79 million increase in added value. Alternatively, when d_{it} is at the first quartile of the sample (and m_{it} is at its mean value), the output elasticity of knowledge flows is 0.01395 ($p < 0.05$), and a 1% increase of knowledge flow leads to a \$0.77 million increase in added value.

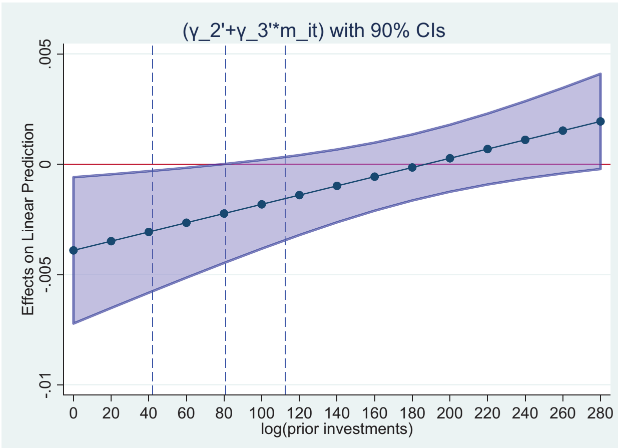
To illustrate the joint effect of prior investments and difficulty of learning on the returns of knowledge flow visually, we present a two-way contour plot in Figure 3. As shown in the figure, the highest return to spillovers accrues to firms that have made significant prior related investments and obtained knowledge flows that are novel (the upper right-hand corner of the figure). In contrast, firms that acquired novel knowledge flow without making prior related investments received the lowest returns (the lower right-hand corner). Indeed, whereas the sign of the first-order effect of spillovers is generally positive, the effect can become negative for low values of prior investments and high values of difficulty of learning. More importantly, it

Figure 1. (Color online) $(\boldsymbol{\gamma}_1' + \boldsymbol{\gamma}_3{}' \boldsymbol{d}_{it})$ at DifferentLevels of $\boldsymbol{d}_{it}$



Notes. Based upon column (5) of Table 3. Vertical dashed lines represent the first, second, and third quartiles of d_{it}.

Figure 2. (Color online) $(\gamma_2' + \gamma_3' m_{it})$ at Different Levels of m_{it}



Notes. Based upon column (5) of Table 3. Vertical dashed lines represent the first, second, and third quartiles of m_{it}.

is evident from the figure that variations in prior related investments lead to drastic changes in the return of spillovers when knowledge is difficult to learn, but they result in only moderate changes in the return of spillovers when knowledge is easy to learn. This can be seen by comparing the variations in colors on the right- and left-hand sides of the figure.

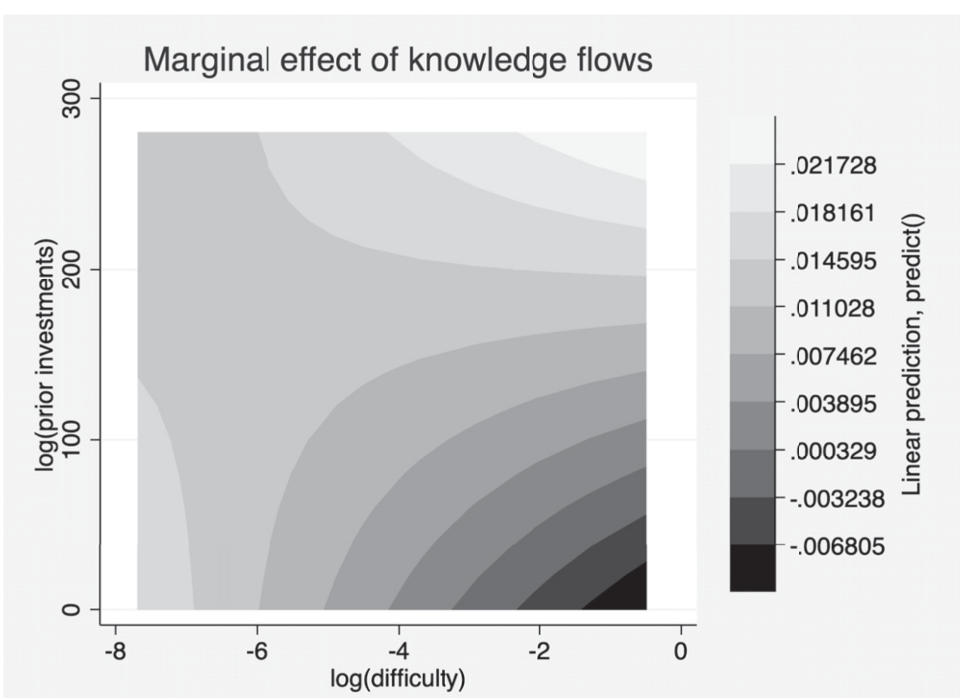
We next discuss some additional aspects of the point estimates of other terms in our model. We note that our estimate of the output elasticity of IT capital (c_{it}, 0.018 in column (5)) is comparable to that in prior literature that uses similar data, such as in Tambe and Hitt (2014b; 0.027) and Tambe and Hitt (2012; 0.032). The differences may be explained by the different sample we use, which consists exclusively of SAP enterprise software adopters.

We further note that the coefficient on our knowledge flow variable (s_{it}) in columns (2)–(5) is not statistically significant; however, it cannot be interpreted directly because of the presence of its interactions with other variables in the model. Finally, it is worth noting that, whereas the coefficient on the variable m_{it} is not statistically significant, this is likely because of two reasons. First, our measure of prior related investments, similar to other work in the ACAP literature that is based on survey measures,[17] has limited variation over time during our sample and, therefore, is difficult to separately identify in a fixed effects model. Second, the time-varying component of m_{it}, based on the number of IT employees, is highly correlated with the included measure of IT capital, c_{it} (correlation coefficient of 0.54, $p < 0.001$), likely causing inflation in the estimate of its standard error.

5.2. Robustness

Our empirical approach of combining the model of ACAP with a production function framework results

Figure 3. Contour Plot, Marginal Effect of Knowledge Flow on Output



in multiple testable implications, and we explore whether the evidence is consistent with the theory. Whereas it is possible that unobserved heterogeneity could influence our estimates, our exploration of multiple testable implications circumscribes the way in which unobserved heterogeneity must influence our results to support alternative explanations. For example, firms that ask questions related to newer enterprise software modules may be systematically different in some ways. However, for these differences to explain our results, their effects must also be weaker in the presence of prior investments. In that way, our combination of the use of the production function approach, the fixed effects panel data, and the exploration of interactions between quasi-fixed (prior module investments) and time-varying (spillover) factors of production makes our empirical approach similar to recent explorations of the effects of complementarities between IT and other production inputs within the IS literature (Aral et al. 2012, Tambe et al. 2012, Wu et al. 2020). Nevertheless, we further present a collage of evidence showing the robustness of our results.

5.2.1. Selection Bias. So far, we have investigated the role of knowledge flows specific to SAP enterprise software on productivity. Arguably, our results could be biased if firms are, at the same time, active in other knowledge forums related to enterprise software, resulting in knowledge flows unobserved to us. For example, this may happen if some firms in our sample have installed enterprise software from another major vendor—such as Oracle—and were active in related forums over the sample period. To investigate the extent to which this influences our findings, we collected data on investments in enterprise software from Oracle made by firms in our sample using the Computer Intelligence database. Using this information, we study whether the effect of SAP-related spillovers and the moderating effects of ACAP variables are significantly different for firms that implemented both systems (SAP and Oracle). We find that our parameter estimates and hypothesis tests are qualitatively similar when we add these controls (results are available upon request).

It is also possible that our estimate of the effect of knowledge inflows is correlated with a selection effect resulting from the firms' endogenous choice of participation in the SCN. For example, if the only firms that choose to seek human capital accumulation through the online forums are those that are more capable of utilizing external knowledge, the positive effect of knowledge inflows on productivity in the population may not be as large as we estimate. To address this selection concern, in column (6) of Tables 3 and 4, we present a subsample analysis in which we use only the firms that eventually received some positive knowledge inflows—that is, the firms that employed the SCN as a mechanism of acquiring human capital—as the sample. This results in a reduced sample with 275 firm-year observations and 58 firms. We note that the coefficient estimates of the terms related to knowledge flows are very similar to those from the full sample (although the significance levels of Hypotheses 2 and 3 drop because of the smaller sample size), alleviating concerns about the implications of such a selection effect.

Finally, we note that, to the extent that selection bias is driven by firm characteristics that do not vary substantially over a relatively short period of time, the employed fixed-effects model reduces selection bias by eliminating all between-firm variation, producing estimates of ACAP variables and interactions that difference out average effects within firms over time.

5.2.2. Alternative Measures. We experiment with alternative measures for some of our key variables (with results reported in Tables 5 and 6). Our baseline measure of m_{it} is based upon the number of SAP modules weighted by the number of IT employees within the firm. Because Table 1 shows that there is cross-firm variance in firm size in our sample, we evaluate two alternative measures of m_{it} that are less directly influenced by firm size. First, we test a model that uses a binary measure of m_{it} (defined using its sample mean) and present the results in column (1) of Table 5. Second, as discussed earlier in Section 4.2.2, we use an alternative measure of prior related investments—a binary indicator that is set to one if the firm's cumulative contribution to SCN prior to year t is greater than the sample mean. In column (2) of Table 5, we present results using this variable. This measure varies to a greater extent within firm and over time when compared with our baseline measure and that of column (1) of Table 5. It is also less correlated with c_{it} (correlation coefficient 0.04, $p = 0.18$) in our sample and so shows a positive and significant coefficient of φ (the coefficient on m_{it}). We present the formal hypothesis tests in Table 6, which shows that both of these measures demonstrate patterns consistent with the absorptive capacity model.

We then present results using a different measure of the difficulty of learning based on the percentage of knowledge flow obtained from SCN forums that are related to business functions. The direction and significance of the results are similar to those in our baseline regressions although the support for Hypotheses 2 and 3 is slightly short of significance at conventional levels. Overall, our empirical tests of the full model lend support to our hypotheses.

In the online appendix, we further examine the robustness of our findings to the mismeasurement of our spillover/knowledge flow variable (presented in

Table 5. Alternative Measures

Variables	(1) Binary measure of *high* m_{it} based on enterprise software investments	(2) m_{it} = high human capital based on forum contributions	(3) d_{it} = log(percentage of business knowledge inflows)
k_{it}	0.11199**	0.10528**	0.11336**
	(0.04397)	(0.04352)	(0.04442)
c_{it}	0.01110**	0.00999*	0.01865**
	(0.00552)	(0.00591)	(0.00736)
l_{it}	0.72815***	0.73125***	0.72790***
	(0.05890)	(0.05490)	(0.05899)
m_{it}	−0.01731	0.06904**	−0.00066
	(0.02374)	(0.02876)	(0.00070)
s_{it}	−0.01479*	−0.00170	0.00073
	(0.00784)	(0.00740)	(0.00889)
$s_{it} \times m_{it}$	0.03514***	0.01791**	0.00010*
	(0.00814)	(0.00881)	(0.00005)
$s_{it} \times d_{it}$	−0.00516***	−0.00371**	−0.00403
	(0.00163)	(0.00157)	(0.00259)
$s_{it} \times d_{it} \times m_{it}$	0.00520**	0.00456**	0.00002
	(0.00204)	(0.00218)	(0.00002)
Log(registered users)	−0.01287	−0.01349	−0.01361
	(0.01578)	(0.01536)	(0.01605)
Log(questions)	−0.01355	−0.01950	−0.01192
	(0.02046)	(0.02047)	(0.02066)
Log(learning by reading)	0.00116	0.00150	0.00100
	(0.00414)	(0.00413)	(0.00416)
Constant	1.65505***	1.68490***	1.66617***
	(0.45959)	(0.44101)	(0.45918)
Year fixed effects	Yes	Yes	Yes
Observations	1,240	1,240	1,240
R^2	0.57951	0.58277	0.57845
Number of firms	275	275	275

Notes. Unless otherwise noted, k_{it} = log(non-IT capital), c_{it} = log(IT capital), l_{it} = log(non-IT labor), m_{it} = log(prior related investment), s_{it} = log(knowledge flows), d_{it} = log(difficulty of learning). The dependent variable is the natural logarithm of value added. All models use firm-level fixed effects and year dummies. Robust standard errors (clustered by firm) are in parentheses. All R^2 values are "within" estimates that do not include the explanatory power of the fixed effects.

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

Table A.3) and show that the findings are robust to different ways of measuring the spillover variable.

5.2.3. Instrumental Variables Regression. As discussed earlier, the amount of knowledge inflow is based upon (1) the number of questions asked and (2) the likelihood of having a question answered. By including the number of questions in our regressions, we control for omitted factors that could be correlated with both the propensity to ask questions and productivity. However, it remains possible that our estimates of knowledge inflows are biased because of omitted factors that influence the likelihood of responses and are correlated with productivity. For example, workers with greater IT skills may have a better reputation in the community and may be better able to articulate their questions, leading to a higher likelihood that their questions are answered.

Table 6. Hypotheses Testing

Hypothesis	(1)	(2)	(3)
Hypothesis 1: $(\gamma_1' + \gamma_3{}' d_{it}) > 0$ with d_{it} at mean	0.03450	0.01679	0.00009
	(p = 0.000)	(p = 0.008)	(p = 0.051)
Hypothesis 2: $(\gamma_2' + \gamma_3{}' m_{it}) < 0)$ with m_{it} at mean	−0.00294	−0.00337	−0.00247
	(p = 0.007)	(p = 0.021)	(p = 0.171)
Hypothesis 3: $\gamma_3{}' > 0$	0.00520	0.00456	0.00002
	(p = 0.011)	(p = 0.037)	(p = 0.211)

Note. p-values are based on two-tailed tests of the null hypothesis that the linear combination of the parameters is zero against the null that is different than zero.

We construct two instrumental variables for knowledge flows to address the latter concern. The first instrument (*IV1*) uses characteristics of the forums in which the firm participates. Some forums have a systematically higher or lower probability of answering a given question. For each question asked by the focal firm, we compute the predicted amount of knowledge inflow based on regressions in which the predictors capture forum-wide characteristics. These characteristics include forum and year fixed effects and also include, for each focal forum-year, the number of questions posted, the number of users, the average number of replies per question, the average number of views, and the average solution rate. All these forum-year variables exclude the focal question. We then sum these predicted values across all the questions asked by the firm in the year and use this as an instrument for knowledge inflows.

The second instrumental variable (*IV2*) takes advantage of exogenous events that draw greater attention to questions raised in some forum-years than those in others. Every year, SAP hosts its largest global business technology event—a conference called Sapphire Now—for its users and partners, offering three full days of networking, strategy, discussions, and education on the latest breakthrough solutions from SAP. In addition to Sapphire Now, SAP also hosts an annual technology education event, SAP TechEd, which offers technical lectures, hands-on workshops, networking opportunities, and SAP executive keynotes covering topics related to the latest developments in SAP products and services. For these two annual events, we collected information related to the most important technical topics (e.g., some key topics in recent years include in-memory computing, big data and real-time analytics, and cloud management) from the conferences' archival websites. We compiled the conference theme topics from product road maps, announcements, keynote speeches, lectures, and workshops at the conferences. We then map these topics at the conferences to the topic forums on SCN.

The instrumental variable is constructed by counting the number of questions raised by the focal firm-year in forums associated with key conference themes in the same year. Questions raised in topic forums that are associated with the key themes at the conferences are more likely to be answered because of the exogenous shift in attention from the community. This could happen via a number of mechanisms: for example, to accelerate the adoption of product or service offerings it promotes at the conferences, SAP may systematically allocate more resources to the technical support of these technologies, some of which would manifest in SCN forums. In addition, conference attendees may socialize with employees from other firms at networking events and workshops and, therefore, strengthen their personal bonds, increasing the likelihood that the questions they raise on conference-related SCN forums get answered.

Because our regression specification in Equation (6) involves interaction terms in which knowledge flow is a component, we need to instrument for these interactions as well because they may also be endogenous. Therefore, in addition to the two instrumental variables *IV1* and *IV2*, we further add the interactions between the two IVs and variables associated with absorptive capacity. In other words, in the IV regression, we have four endogenous variables: s_{it}, $m_{it}s_{it}$, $d_{it}s_{it}$, and $m_{it}d_{it}s_{it}$, and we use eight instrumental variables for them: $IV1_{it}$, $m_{it}IV1_{it}$, $d_{it}IV1_{it}$, $m_{it}d_{it}IV1_{it}$, $IV2_{it}$, $m_{it}IV2_{it}$, $d_{it}IV2_{it}$, and $m_{it}d_{it}IV2_{it}$.

We report the results of the two-stage least squares (2SLS) model, together with a summary of first-stage regressions in column (1) of Table 7. For all four endogenous variables, the Angrist–Pischke first stage *F*-tests of exclusion restrictions reject the null, suggesting that the instruments are not weak. The Stock–Yogo critical values further confirm the validity of the instruments (in all cases the Cragg–Donald Wald *F* statistics are greater than the critical values, and they are also greater than the rule-of-thumb value of 10).[18] In addition, in the overidentification test, the Hansen *J* statistic has a value of 7.39, which cannot reject the null that the set of instruments are valid.

We observe that, in column (1), the coefficient estimates of $m_{it}s_{it}$, $d_{it}s_{it}$, and $m_{it}d_{it}s_{it}$ as shown in the second-stage IV regression are very similar and in the same direction as those obtained from our baseline regression in column (5) of Table 3 (0.00013 versus 0.00015, −0.00318 versus −0.00404, and 0.00002 versus 0.00002, respectively). This is confirmed by a Hausman test comparing the baseline model and the IV regression, which cannot reject the null that the difference in the coefficient estimates is not systematic ($\chi^2(12)$ = 0.88, $p > 0.10$). We further present the formal hypothesis tests based on the IV regression results in Table 8. Again, the values of the test statistics for Hypotheses 1–3 are very similar in their magnitude to those based on our baseline regression in column (5) of Table 4 (0.00013 versus 0.00014, −0.00126 versus −0.00233, and 0.00002 versus 0.00002 for Hypotheses 1–3, respectively) although the significance levels drop because of the inflation in the estimated values of standard errors, which is not uncommon when instrumenting for multiple endogenous interaction terms.

We extend our baseline IV analysis in two ways. First, we reestimate our baseline IV model when replacing the continuous m_{it} with its binary counterpart (with the uninstrumented model as in column (1) of Table 5). In this model, the increase in the estimated values of the standard errors is not as severe, and all three hypotheses are supported at the conventional

Table 7. Instrumenting for Knowledge Inflows

	(1)	(2)	(3)
Corresponding uninstrumented model	Column (5) of Table 3	Column (1) of Table 5	Column (5) of Table 3
Variables	Second stage of 2SLS	Second stage of 2SLS Binary measure of *high* m_{it}	System GMM
k_{it}	0.11365***	0.11196***	−0.0493*
	(0.03723)	(0.03702)	(0.0266)
c_{it}	0.01914***	0.01153**	−0.0127
	(0.00652)	(0.00490)	(0.0093)
l_{it}	0.72839***	0.72842***	−0.0747
	(0.04992)	(0.05011)	(0.0539)
m_{it}	−0.00065	−0.01703	−0.0002
	(0.00059)	(0.02090)	(0.0004)
s_{it}	0.00445	−0.01050	−0.0450*
	(0.01758)	(0.01111)	(0.0232)
$s_{it} \times m_{it}$	0.00013	0.03618***	0.0003**
	(0.00009)	(0.01126)	(0.0002)
$s_{it} \times d_{it}$	−0.00318	−0.00496***	−0.0115**
	(0.00326)	(0.00182)	(0.0047)
$s_{it} \times d_{it} \times m_{it}$	0.00002	0.00667***	0.0001**
	(0.00002)	(0.00253)	(0.0000)
Log(registered users)	−0.00972	−0.01104	0.0161
	(0.01475)	(0.01449)	(0.0476)
Log(questions)	−0.02299	−0.01567	0.0045
	(0.02269)	(0.02111)	(0.0445)
Log(learning by reading)	0.00201	0.00124	−0.0039
	(0.00397)	(0.00386)	(0.0121)
Lag of log(added value)			1.1433***
			(0.0815)
Constant			−0.0861
			(0.0880)
Year fixed effects	Yes	Yes	Yes
Observations	1,227	1,227	974
Number of firms	262	262	266
Cragg–Donald Wald *F*-statistic	39.294	45.682	
Hansen *J* statistic	4.991 ($p > 0.10$)	4.781 ($p > 0.10$)	
R^2	0.57786	0.57925	
Autocorrelation test, order 1			z = −3.46 ($p = 0.001$)
Autocorrelation test, order 2			z = −0.98 ($p = 0.329$)

Notes. Unless otherwise noted, k_{it} = log(non-IT capital), c_{it} = log(IT capital), l_{it} = log(non-IT labor), m_{it} = log(prior related investment), s_{it} = log(knowledge flows), d_{it} = log(difficulty of learning). The dependent variable is the natural logarithm of value added. All models use firm-level fixed effects and year dummies. Robust standard errors (clustered by firm) are in parentheses. All R^2 values are "within" estimates that do not include the explanatory power of the fixed effects. Thirteen observations were dropped in the IV regressions (columns (1) and (2)) because of singletons.

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

significance levels ($p < 0.01$, $p < 0.1$, and $p < 0.01$, respectively) as we show in column (2) of Table 8.

Second, we have conducted a system generalized method of moments (GMM) estimation that incorporates the estimation approach attributed to Arellano and Bond (1991) and Blundell and Bond (1998). Our approach follows the methodological suggestions provided by Roodman (2009), using the *xtabond2* GMM estimator in Stata. In particular, the dynamic panel estimator includes one lag of the dependent variable, which, together with knowledge inflows and its interactions with the ACAP terms (i.e., $s_{it}, m_{it}s_{it}, d_{it}s_{it}$, and $m_{it}d_{it}s_{it}$), is treated as endogenous. We use deeper lags of the dependent variable and endogenous variables as GMM-style instruments. In addition, we include $IV1_{it}$, $m_{it}IV1_{it}$, $d_{it}IV1_{it}$, $m_{it}d_{it}IV1_{it}$, $IV2_{it}$, $\mathrm{m}_{it}IV2_{it}$, $d_{it}IV2_{it}$, and $m_{it}d_{it}IV2_{it}$ as standard IV-style instruments in the levels equation. We present the estimation results in column (3) of Table 7 and, in keeping with presentation of these models, present the autocorrelation tests of orders 1 and 2. We note that these latter statistics are consistent with acceptable values in these GMM models: rejecting the null of no autocorrelation of order 1 but finding no evidence of autocorrelation of order 2 (for example, see Roodman 2009, Nagle 2019). The results are qualitatively consistent with our benchmark results, and the two-tailed tests provide strong statistical significance ($p < 0.05$ for all three hypotheses).

6. Conclusions

This paper shows that the productivity effects of knowledge flows related to the implementation and

Table 8. Hypotheses Testing

Hypothesis	(1)	(2)	(3)
Hypothesis 1: $(\gamma_1' + \gamma_3' d_{it}) > 0$ with d_{it} at mean	0.00013	0.03535	0.00032
	(p = 0.140)	(p = 0.001)	(p = 0.048)
Hypothesis 2: $(\gamma_2' + \gamma_3' m_{it}) < 0)$ with m_{it} at mean	−0.00126	−0.00212	−0.00544
	(p = 0.549)	(p = 0.092)	(p = 0.045)
Hypothesis 3: $\gamma_3' > 0$	0.00002	0.00667	0.00007
	(p = 0.163)	(p = 0.008)	(p = 0.041)

Note. p-values are based on two-tailed tests of the null hypothesis that the linear combination of the parameters is zero against the null that is different than zero.

use of enterprise software are critically moderated by a firm's prior IT investments, the nature of external IT knowledge flows, and their interaction. In this way, our findings extend implications of ACAP theory from the R&D literature to a new setting. We also provide boundary conditions for when prior results applying ACAP to business process innovation in the IS setting do not hold.

We adopt a novel measurement strategy that allows us to examine activity in an online discussion forum, a channel increasingly used by firms to augment the human capital necessary to deploy IT systems. By combining a novel data source with an established theoretical framework, we show that the effect of external knowledge flows is stronger for firms with prior investments in enterprise software and lower when external IT knowledge is difficult to learn, such as for knowledge originating from relatively newer and emerging discussion forums. However, it is precisely in these environments that prior investments in enterprise software have their most significant impact on facilitating the absorption of knowledge and, thus, increasing a firm's productivity.

We contribute to the existing literature on IT spillovers by fully applying the essence of the ACAP theory to the context of enterprise software. We show that, on the one hand, IT spillovers related to enterprise software are not "free," and only firms with significant prior ERP investments can benefit from them. On the other hand, failure to consider the dual effects of enterprise IT investments leads to underestimation of their true returns. Whereas prior IS literature recognizes the importance of a firm's ACAP as a key capability moderating knowledge transfer and productivity, it mainly focuses on the path-dependent component of ACAP—the stock of prior IT knowledge as a function of prior IT investments (Roberts et al. 2012). We contribute to the IS literature by focusing on an aspect of ACAP theory that has been neglected in the past—features of knowledge that affect the difficulty of learning. Given the differences in the manner in which knowledge is absorbed between the R&D and IS settings, it is unclear ex ante whether prior results from the R&D literature hold, and prior literature is unable to inform this gap in the literature because of the challenges in measuring the nature of knowledge in the IS setting. Our measurement strategy on knowledge flows allows us an unusual opportunity to measure the interrelated effects of knowledge spillovers, prior related investments, and the type of knowledge.

Our results have several managerial implications. First, firms that fail to account for the indirect effects of their IT investments likely underestimate their productivity implications. For example, the shift away from on-premises to cloud computing creates broader implications for firms. Historically, investments in applications software were accompanied by investments in how to deploy the systems. That is, firms deploying enterprise software were required to make complementary investments in business process innovation. As firms increasingly deploy service-based application software that may require smaller investments to deploy, this may influence their ability to respond to new enterprise IT-based opportunities in the future. This shares some similarities with earlier concerns about whether "offshoring" software development would lead to a hollowing out of the labor force in the United States (e.g., Levy and Murnane 2005).

Further, the circumstances that we identify in which prior investments are most valuable—for knowledge related to new applications—are precisely those in which firms are most likely to seek external knowledge because of a lack of codified best practices. That is, our results highlight how firms without prior related knowledge are likely to struggle in deploying frontier applications.

Our results also inform understanding of how firms interact with and gain value from an increasingly important source of knowledge: online communities. In particular, they provide insights on why firms participate in online communities, such as SCN. Some studies argue that workers contribute to open source projects to develop their skills (e.g., Lakhani and von Hippel 2003, Lakhani and Wolf 2005), and more recent work contends that firms provide incentives for workers to contribute to such projects to accumulate the human capital there (Mehra et al. 2011). Consistent with the work of Nagle (2018), our results suggest the existence of benefits of participation not only through inflows but also through contributions because making contributions results in the accumulation of related IT knowledge, which, in turn, increases absorptive capacity.

Although our research contributes to the literature on IT spillovers, it is noteworthy that the process by which spillovers are generated in our context differs significantly. In contrast to earlier work on IT spillovers (e.g., Cheng and Nault 2007, 2012; Tambe and Hitt 2014a, b) and the traditional R&D literature on absorptive capacity (e.g., Cohen and Levinthal 1989, 1990), knowledge flows in our setting are not externalities arising from investments in product or business process innovation from firms in the same industry, supply chain, or network. Instead, they arise from deliberate decisions by employees of firms to ask and answer questions and, in that way, bear some similarity to the nature of knowledge flows arising from the transactional relationship between an IT services provider and its clients (Chang and Gurbaxani 2012a, b). Given that the nature of the knowledge transferred is likely already customized to the firm's needs to some degree, the continued importance of absorptive capacity is striking.

Our research approach offers a new means of measuring the content of flows of knowledge between firms. As noted elsewhere, these have been difficult to measure in the past. It is useful, however, to characterize the differences between our approach and prior papers that have used the ACAP framework in the IS literature, explicating the advantages and disadvantages of each. Prior work primarily uses surveys to measure constructs in the ACAP model (Roberts et al. 2012). For example, some researchers use surveys to measure ACAP as an asset (Ko et al. 2005, Xu and Ma 2008) or as a capability (Armstrong and Sambamurthy 1999, Pavlou and El Sawy 2006), emphasizing the role of human capital of the firm in facilitating the absorption of external knowledge. One challenge faced by many of these papers is that they are often costly to implement and suffer from nonresponse or recall biases. In contrast, whereas we have unusually direct measures of knowledge flows and its characteristics, we capture heterogeneity in the ability of firms to absorb new knowledge indirectly through prior related investments (for another recent paper that uses this approach, see Trantopoulos et al. 2017). We further note that, in contrast to prior work, which seems to measure heterogeneity in knowledge absorption in general, our approach focuses on knowledge flow and its absorption within a specific online community related to the use of enterprise software. However, as we note elsewhere, such forums are becoming an increasingly important way of transferring knowledge across IT workers (Howe 2008, Boudreau and Lakhani 2009).

Our work also contributes to prior research on the interrelationships between IT investment, business process innovation, and productivity.[19] The literature on business process innovation has been hampered by a number of challenges, namely, the difficulty of measuring the inputs and outputs of the innovation process. Although measurement of innovation is always problematic (Mortensen and Bloch 2005, Cohen 2010), measurement of business process innovation is particularly difficult because it leaves behind no tangible "footprints," such as patents in the R&D literature. Our work provides further insights on the role of external knowledge flows in augmenting internal human capital through a unique measurement strategy that uses online behavior to capture inputs into business process innovation that could not previously be measured directly. Whereas we acknowledge that our measures may not capture all such human capital accumulation, they are in the spirit of the literature on business process innovation that uses proxies for hard-to-measure inputs and outputs and acknowledges that output elasticities may capture variance related to some kinds of unmeasured activity (e.g., Bresnahan et al. 2002, Anderson et al. 2003). In particular, we wait for and encourage further work that may find alternative strategies of measuring the key inputs into the ACAP model.

As in any study seeking to measure the productivity implications of IT investments, business process innovation, and human capital acquisition, our study has some limitations. One advantage of our study over prior work is our ability to measure knowledge flows using archival data. However, as noted elsewhere, our estimation strategy and robustness are shaped by the unique data-generating process of knowledge acquisition in our setting, which involves endogenous choices to raise and answer questions. Further, as in other studies, we must be cognizant of whether organizational investments, such as external knowledge acquisition are correlated with other unobserved variables. To address these concerns, we examined the robustness of our results to a range of alternative strategies. Although our results are robust to these efforts, we leave it to future work to study the robustness of our results to other contexts.

Our research highlights opportunities for new research. For example, one interesting possibility is investigating, at a more disaggregated level, how related human capital investments influence the benefits that accrue to individual workers from participation in related communities (Huang and Zhang 2016). This can be accomplished by tying community activity to databases of worker skills, using data from sites such as LinkedIn. We hope our research spurs additional work in these important areas.

Acknowledgments
The authors thank Erik Brynjolfsson, Lorin Hitt, Paul Hofmann, Frank Nagle, Alex Oettl, Marshall Van Alstyne, and seminar participants at Arizona State University, Boston University, Columbia Business School, Ecole Polytechnique Fédérale de Lausanne (Switzerland), Emory University, Georgia State University, Harvard Business School, IESE (Spain), Lehigh University, Massachusetts Institute of Technology, New York University, Stanford University, Université Libre de Bruxelles (Belgium), University of Bremen (Germany), University of California, Irvine, University of Rochester, University of Illinois Urbana-Champaign, Academy of Management Paper Symposium on Information Technology Spillovers (2013), National Bureau of Economic Research Summer Institute (2013), Organization Science Winter Conference (2018), and Workshop on Information Systems and Economics (2012) for helpful comments and suggestions. The authors thank Bin Li for his help collecting the SAP Community Network (SCN) data. They also thank Michael Zhang (the senior editor), the associate editor, and the reviewers for their constructive comments and insights. All opinions and errors are the authors' own.

Appendix 1. Details on Computation of IT Capital
As noted in the paper, our measure of IT capital is derived from the CI database. We adopt the method used by Brynjolfsson and Hitt (1995), Dewan and Min (1997), Gu et al. (2008), and Hitt and Brynjolfsson (1996) and measure the IT capital stock using an estimate of the market value of the IT hardware systems plus three times the current year IT labor expenses. The first component of this variable is equal to the market value of total PCs and servers currently owned by the firm, converted to constant 2005 U.S. dollars. To be specific, we collect market prices of PCs and servers in the United States from a two-report series produced by the Gartner Dataquest Market Statistics database—Gartner Worldwide Server Forecast and Gartner Worldwide PC Forecast—from 2004 to 2008. This two-report series presents detailed statistics on the number of shipments, prices, vendor revenues, and other related information about PC and servers broken down to the level of each geographic region and market segment.[20] Our market prices for PCs and servers are calculated as the average user price across their respective market segments within the United States. These prices are then multiplied by the quantities of PCs and servers owned by the firm to derive the market value of the IT computer assets for each firm. Finally, we deflate the market value by the Bureau of Economic Analysis price index for computers and peripheral equipment.

The second component of IT capital stock is IT-related labor expenses. The CI database provides the number of IT employees of the sample firms at the site level (which falls into one of the following ranges: 1–4, 5–9, 10–24, 25–49, 50–99, 100–249, 250–499, and 500 or more), and a site represents a particular firm location, much like the concept of establishment in census data. We aggregate the site-level employee numbers to the firm level to derive the total number of IT-related employees hired by the firm. For each range, we take the middle value of the range as the number of IT employees. IT labor prices are obtained from the occupational employment and wage estimates series from the BLS OES, and we use the mean annual wage of computer and mathematical occupations as the average labor price for IT employees. As the wage reported by the OES series does not reflect benefits, we multiply the wage number by the ratio of total compensation to salary, which is obtained from BLS ECEC series. The IT labor expense is then deflated by the BLS ECI for private industry workers.

Appendix 2. Additional Robustness Tests of the Absorptive Capacity Model
Although we use fixed effects to control for time-invariant firm-level heterogeneities, some time-varying, unobserved factors might be correlated with both our measurement of knowledge spillovers and firm productivity. In addition, the presence of potential measurement errors in our spillover variable may lead to bias in estimation. Although it is impossible to control for all the unobservable factors, in the following, we present a systematic discussion of the remaining endogeneity issues and the measures we take to address each of them. We focus on two sources of potential endogeneity in this discussion: (1) mismeasurement of our spillover variable and (2) IT spillovers through investments made by firms in the same industry that may be correlated with our measure.

Table A.1. Evolution of SAP Community Network

Year	Number of registered users	Number of active forums	Number of new threads initiated in the year	Average number of replies for threads initiated in the year	Fraction of questions solved	Fraction received helpful answers	Fraction received very helpful answers	Number of days until correct answer
2004	19,289	57	16,296	4.679	0.107	0.073	0.098	13.378
2005	43,226	83	67,225	5.394	0.244	0.271	0.295	4.735
2006	80,981	141	176,422	5.160	0.242	0.293	0.314	3.359
2007	137,552	179	394,183	4.731	0.227	0.260	0.287	4.219
2008	198,975	209	463,740	4.625	0.252	0.255	0.256	4.512

Table A.2. Industry Segments of the Sample

Two-digit NAICS	Description	Frequency	Percentage
11	Agriculture, forestry, fishing, and hunting	5	0.4
21	Mining, quarrying, and oil and gas extraction	25	2.02
22	Utilities	104	8.39
23	Construction	8	0.65
31–33	Manufacturing	824	66.45
42	Wholesale trade	51	4.11
44–45	Retail trade	48	3.87
48–49	Transportation and warehousing	17	1.37
51	Information	71	5.73
52	Finance and insurance	12	0.97
53	Real estate and rental and leasing	10	0.81
54	Professional, scientific, and technical services	31	2.5
56	Administrative and support and waste management and remediation services	10	0.81
62	Healthcare and social assistance	9	0.73
72	Accommodation and food services	15	1.21
Total		1,240	100

Table A.3. Robustness Tests of Absorptive Capacity Model

Variables	(1) Alternative measures of spillovers	(2)	(3) Industry Spillover Pool	(4)
k_{it}	0.11383**	0.11367**	0.11411**	0.10994**
	(0.04428)	(0.04424)	(0.04419)	(0.04497)
c_{it}	0.01811**	0.01820**	0.01735**	0.01819**
	(0.00737)	(0.00737)	(0.00748)	(0.00743)
l_{it}	0.72697***	0.72712***	0.72737***	0.72923***
	(0.05877)	(0.05873)	(0.05879)	(0.05887)
m_{it}	−0.00064	−0.00063	−0.00061	−0.00065
	(0.00070)	(0.00070)	(0.00070)	(0.00070)
s_{it}	−0.00834	−0.01069	−0.00822	−0.00840
	(0.01032)	(0.01847)	(0.01013)	(0.01011)
$s_{it} \times m_{it}$	0.00015***	0.00025***	0.00015***	0.00015***
	(0.00005)	(0.00009)	(0.00005)	(0.00006)
$s_{it} \times d_{it}$	−0.00417**	−0.00660**	−0.00410**	−0.00409**
	(0.00195)	(0.00316)	(0.00193)	(0.00188)
$s_{it} \times d_{it} \times m_{it}$	0.00002**	0.00004**	0.00002**	0.00002**
	(0.00001)	(0.00002)	(0.00001)	(0.00001)
Industry spillover pool			0.01855	0.01130
			(0.01395)	(0.01122)
Log(registered users)	−0.01368	−0.01182	−0.01330	−0.01388
	(0.01593)	(0.01603)	(0.01601)	(0.01598)
Log(questions)	−0.01306	−0.01480	−0.01296	−0.01264
	(0.02060)	(0.02131)	(0.02063)	(0.02053)
Log(learning by reading)	0.00119	0.00132	0.00106	0.00105
	(0.00418)	(0.00424)	(0.00420)	(0.00413)
Constant	1.66753***	1.66635***	1.53008***	1.61534***
	(0.45824)	(0.45838)	(0.45217)	(0.45989)
Observations	1,240	1,240	1,240	1,240
Number of firms	275	275	275	275
R^2	0.57865	0.57854	0.57911	0.57909

Notes. The dependent variable is the natural logarithm of value added. All models use firm-level fixed effects and year dummies. Robust standard errors (clustered by firm) are in parentheses. All R^2 values are "within" estimates that do not include the explanatory power of the fixed effects. In column (1), the spillover measure includes only reward points from correct and very helpful answers. In column (2), the spillover measure is defined as the number of resolved questions not weighted using reward points. In column (3), the spillover pool is calculated based on the IT investments of firms in the same three-digit NAICS industry. In column (4), the spillover pool is calculated based on the IT investments by firms in the same three-digit NAICS industry and had SAP enterprise software installations.

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

Table A.4. Hypotheses Testing

Hypothesis	(1)	(2)	(3)	(4)
Hypothesis 1: $(\gamma_1' + \gamma_3' d_{it}) > 0$ with d_{it} at mean	0.00015	0.00025	0.00014	0.00015
	(p = 0.006)	(p = 0.008)	(p = 0.007)	(p = 0.007)
Hypothesis 2: $(\gamma_2' + \gamma_3' m_{it}) < 0)$ with m_{it} at mean	−0.00416	−0.00660	−0.00236	−0.00232
	(p = 0.033)	(p = 0.038)	(p = 0.069)	(p = 0.067)
Hypothesis 3: $\gamma_3' > 0$	0.00002	0.00004	0.00002	0.00002
	(p = 0.034)	(p = 0.022)	(p = 0.033)	(p = 0.029)

Note. p-values are based on two-tailed tests of the null hypothesis that the linear combination of the parameters is zero against the null that is different than zero.

First, endogeneity concerns may arise because of the way we measure the IT spillover variable. For example, one potential issue with our baseline spillover measure is that knowledge seekers may not pay enough attention to the answers posted or simply lack the expertise to judge the quality of the answers, leading to mismeasurement of the spillover variable. Moreover, we might mismeasure the magnitude of spillovers if a knowledge seeker rewards too many knowledge contributors by marking their posts as helpful (because, unlike correct answers and very helpful answers, the number of helpful answers per thread is not limited).

Such measurement errors, if they exist, are most likely to result in an attenuation bias in the estimates. However, to assess how measurement error might influence our results, we perform two separate analyses. First, we calculate the spillover variable using reward points from only correct and very helpful answers, therefore preventing the large number of helpful answers from inflating the spillover variable. We present the result of this analysis in column (1) of Table A.3. Second, we construct the spillover variable by simply counting the number of questions that are resolved (questions that received either a correct answer or at least a very helpful answer) without using reward points as weight. The result is presented in column (2) of Table A.3. We find that the findings are robust to different ways of measuring the spillover variable.

Second, endogeneity may cause bias in our coefficient estimate of knowledge spillover if our measure of spillover is correlated with those that occur via other spillover channels, such as those that are mediated through external knowledge pools. In prior literature, such external knowledge pools are usually modeled as the weighted sum of IT investments made by other firms in the same industry or those facing the same technological opportunities (Tambe and Hitt 2014b). We present two sets of results in which we explicitly account for other spillover channels using a pooled approach, similar to the one adopted by Tambe and Hitt (2014b). First, we construct the spillover pool using industry proximity and define the pool as the sum of IT investments made by all other companies (among the Fortune 1,000) in the same three-digit NAICS industry and present the results in column (3) of Table A.3. Second, we also construct the spillover pool using both industry and technological proximity. Specifically, we define the pool as the sum of the IT investments by all other firms (among the Fortune 1,000) that (1) fall into the same three-digit NAICS industry as the focal firm and (2) had installed SAP enterprise software prior to 2004. The results of incorporating the spillover pool are presented in column (4) of Table A.3. We find that the coefficient estimate remains robust after we add the pooled measures of IT spillovers.[21]

Appendix 3. Comparison with the Measure of Supply Chain Spillovers

Prior research on IT spillovers highlights the distinction between knowledge and rent spillovers; the latter happens when factor inputs are purchased from other industries at a price that does not fully reflect the improvements in the quality resulting from the use of IT (Chang and Gurbaxani 2012a). However, earlier studies are unable to disentangle the two empirically because of measurement difficulties (Mun and Nadiri 2002, Cheng and Nault 2007, Chang and Gurbaxani 2012a). Our novel measurement of knowledge flow, in contrast, offers a unique opportunity to study the two jointly. For comparison purposes, we follow the procedure described in Chang and Gurbaxani (2012b) and compute the IT spillovers embedded in purchases of intermediate inputs from upstream industries (named *SP2* in Chang and Gurbaxani 2012b), using industry level input–output tables published by the Bureau of Labor Statistics.

The purpose of the exercises performed in this section is threefold. First, if the knowledge flow variable we construct captures variation in human capital deepening through the acquisition of external knowledge rather than IT spillovers that are embodied in intermediate inputs, then adding IT spillovers through the supply chain into the regression should have little impact on our findings. We explore this possibility by adding the supply chain IT spillovers (in its log form, named s'_{it}) into our baseline regression and present the results in column (1) of Table A.5. We find that, as expected, the coefficient estimates of $m_{it}s_{it}$, $d_{it}s_{it}$, and $m_{it}d_{it}s_{it}$ are qualitatively similar to our earlier estimates. Interestingly, we observe that, when this pool-based, embedded IT spillover variable is added to the model, it leads to a lower (and insignificant) estimate of the output elasticity of non-IT capital relative to column (5) of Table 3 as well as a lower output elasticity of non-IT labor, suggesting that the measure may be capturing variations beyond IT spillovers. This correlation between IT spillover and other factors of production has been identified in prior research (Tambe and Hitt 2014b).

Table A.5. Horse Race Between Knowledge Flows and Supply Chain Embedded Spillovers

Variables	(1) Knowledge flows	(2) Supply chain spillovers	(3) Both spillovers	(4) SAP-specific prior investment versus general prior IT investment
k_{it}	−0.06188	−0.06180	−0.06122	−0.06192
	(0.04560)	(0.04495)	(0.04542)	(0.04537)
c_{it}	0.01491**	0.01555**	0.01439**	−0.00016
	(0.00707)	(0.00668)	(0.00678)	(0.01344)
l_{it}	0.39245***	0.38915***	0.38773***	0.39003***
	(0.05750)	(0.05771)	(0.05772)	(0.05724)
m_{it}	−0.00122*	−0.00265**	−0.00262**	−0.00114
	(0.00072)	(0.00131)	(0.00131)	(0.00070)
Supply chain spillover (s'_{it})	0.72232***	0.70127***	0.70219***	0.71300***
	(0.09178)	(0.09108)	(0.09157)	(0.09257)
Knowledge flows (s_{it})	−0.00839	0.00940	−0.00937	−0.00807
	(0.00879)	(0.00652)	(0.00991)	(0.00899)
$s_{it} \times m_{it}$	0.00016***		0.00017***	0.00016***
	(0.00005)		(0.00006)	(0.00006)
$s_{it} \times d_{it}$	−0.00287*		−0.00217	−0.00269
	(0.00168)		(0.00183)	(0.00176)
$s_{it} \times d_{it} \times m_{it}$	0.00003**		0.00002	0.00002**
	(0.00001)		(0.00002)	(0.00001)
$s'_{it} \times m_{it}$		0.00034	0.00034	
		(0.00022)	(0.00022)	
$s'_{it} \times d_{it}$		0.00038	−0.00103	
		(0.00193)	(0.00255)	
$s'_{it} \times d_{it} \times m_{it}$		−0.00001	0.00000	
		(0.00001)	(0.00002)	
$s'_{it} \times c_{it}$				0.00321
				(0.00277)
Log(registered users)	−0.00490	−0.00303	−0.00505	−0.00432
	(0.01236)	(0.01272)	(0.01297)	(0.01232)
Log(questions)	−0.00383	−0.00352	−0.00278	−0.00353
	(0.01732)	(0.01756)	(0.01800)	(0.01735)
Log(learning by reading)	−0.00089	−0.00122	−0.00112	−0.00106
	(0.00365)	(0.00370)	(0.00372)	(0.00366)
Constant	2.50685***	2.61181***	2.61385***	2.56042***
	(0.37537)	(0.36727)	(0.36998)	(0.37201)
Year fixed effects	Yes	Yes	Yes	Yes
Observations	1,240	1,240	1,240	1,240
R^2	0.69920	0.69980	0.70056	0.69959
Number of firms	275	275	275	275

Notes. The dependent variable is the natural logarithm of value added. All models use firm-level fixed effects and year dummies. Robust standard errors are in parentheses. All R^2 values are "within" estimates that do not include the explanatory power of the fixed effects. In column (1), we add a measure of supply chain spillovers, but we test for the presence of ACAP using our original measure of knowledge flows. In column (2), we interact supply chain spillovers with m_{it} and d_{it}, leaving knowledge flows s_{it} only as a control. In column (3), we interact both s_{it} and s'_{it} with m_{it} and d_{it}. In column (4), we combine tests for ACAP using s_{it} with an interaction between supply chain spillovers s'_{it} with generic IT capital investment c_{it}.

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

Table A.6. Hypotheses Testing

Hypothesis	(1)	(2)	(3)	(4)
Hypothesis 1: $(\gamma_1' + \gamma_3{}' d_{it}) > 0$ with d_{it} at mean	0.00016	N/A	0.00016	0.00015
	(p = 0.003)		(p = 0.004)	(p = 0.004)
Hypothesis 2: $(\gamma_2' + \gamma_3{}' m_{it}) < 0)$ with m_{it} at mean	−0.00081	N/A	−0.00032	−0.00074
	(p = 0.439)		(p = 0.782)	(p = 0.501)
Hypothesis 3: $\gamma_3{}' > 0$	0.0003	N/A	0.00002	0.00002
	(p = 0.018)		(p = 0.159)	(p = 0.027)

Note. p-values are based on two-tailed tests of the null hypothesis that the linear combination of the parameters is zero against the null that is different than zero.

Second, we use the embedded IT spillover measure to conduct a falsification test. If our model of absorptive capacity is correct, we expect the patterns as predicted in the absorptive capacity model do not apply to IT spillovers through a firm's supply chain when we interact those spillovers with prior investments in enterprise software and our measure of difficulty of learning (based on activity in the forum). In column (2) of Table A.5 we present a model in which we interact s'_{it} with m_{it}, d_{it}, and $m_{it}d_{it}$. We find that the three interaction terms $m_{it}s'_{it}$, $d_{it}s'_{it}$, and $m_{it}d_{it}s'_{it}$ display entirely different patterns than predicted by the ACAP model. We add both sets of interactions in column (3) of Table A.5. We find that the estimates using our earlier measure of knowledge flows s_{it} remains directionally consistent with the ACAP model predictions although the significance of $d_{it}s_{it}$, and $m_{it}d_{it}s_{it}$ drop below conventional levels. In contrast, we do not find estimates using s'_{it} are consistent with the ACAP model.

Third, although we show that the embedded IT spillovers do not interact with enterprise software–specific investments (m_{it}) in a way that is predicted by the ACAP model, we test the conjecture that its contribution to productivity may be stronger with greater *generic* IT investments. Therefore, we interact s'_{it} with a firm's IT capital (c_{it}) and add the interaction into the regression. We find the coefficient estimate of the interaction is directionally consistent with our conjecture but statistically short of being significant at conventional levels. The results are presented in column (4) of Table A.5. In summary, this set of horse races between knowledge flows and IT spillovers through supply chain relationships demonstrates that our knowledge flow variable indeed reflects a different spillover mechanism than that embedded in a firm's purchase of intermediate inputs.

Endnotes

[1] For a review of the literature, see, for example, Brynjolfsson and Milgrom (2012). For a recent example, see Tambe and Hitt (2014a).

[2] Investments in education, training, health, and values that cannot easily be separated from people are regarded as human capital (Becker 2008). As subsequently discussed in further detail, in our setting, knowledge acquisition builds human capital through a variety of formal and informal means. We, therefore, use the terms "acquire knowledge" and "accumulate human capital" interchangeably.

[3] See http://nymag.com/intelligencer/2017/03/the-hidden-power-of-stack-overflow.html.

[4] See https://stackoverflow.blog/2019/01/18/state-of-the-stack-2019-a-year-in-review/.

[5] See https://insights.stackoverflow.com/survey/2019.

[6] See https://en.wikipedia.org/wiki/SAP_Community_Network.

[7] Our model and data analysis assume that knowledge stocks are formed based on the cumulated enterprise IT investments and flows of external knowledge related to enterprise software. We explore the robustness of our results to the inclusion of other sources of spillovers in the empirical section.

[8] Note that, in C&L, $\gamma_0 = 0$. We add it here to be more general, and it allows absorptive capacity to have an independent effect on the production function, as evident in Equations (4) and (5).

[9] This argument is also used to explain higher levels of spatial clustering in the early phases of an industry lifecycle, when new knowledge plays an important role and the associated transfer of tacit knowledge is facilitated by geographic proximity (Audretsch and Feldman 1996).

[10] Retrieved from http://www.bls.gov/mfp/mprdload.htm.

[11] We use the $\log(1+S_{it})$ transformation in all regressions as a measure of s_{it} to avoid loss of observations when $S_{it}=0$. To probe the robustness of our procedure, we reestimate our models adding a dummy variable = 1 when $S_{it}=0$ and find that our results are qualitatively unchanged. Results are available upon request.

[12] For example, among all the discussion threads that are initiated by U.S. knowledge seekers during our sample period, only 48% (23,973 out of 49,977) of them have a seeker who reported an employer.

[13] See https://blogs.sap.com/2008/01/22/business-objects-diamond-bring-us-value/.

[14] Some examples of technical oriented forums are Java Programming, Form Printing, SAP on SQL Server, and Data Transfers. Some examples of business-oriented forums are Logistic Materials Management, Sales and Distribution General, and ERP Operations–Quality Management.

[15] Typical SAP technical modules are ABAP (Advanced Business Application Programming) and BASIS (Business Application Software Integrated Solution). Typical SAP functional modules are FICO (Finance & Controlling), HR (Human Resource), and MM (Material Management).

[16] For additional details on computing linear combinations of coefficients, see Jaccard et al. (1990) and Aiken et al. (1991).

[17] For further discussion, see Roberts et al. (2012).

[18] Stock–Yogo critical values are not reported because of space constraints; they are available from the authors upon request.

[19] See, for example, Bartel et al. (2007), Bresnahan and Greenstein (1996), Bresnahan et al. (2002), Dranove et al. (2014), and Ichniowski and Shaw (2003).

[20] Gartner Dataquest defines PC market segments as desk-based, mobile, professional, and home. Server market segments are defined by CPU types, which include x86, IA64, RISC, and other. The database covers the global regions of Asia/Pacific, Eastern Europe, Latin America, Middle East and Africa, and Western Europe. Several country-level statistics are also available, including the United States, Canada, and Japan.

[21] We note that, unlike Tambe and Hitt (2014b), the spillover pool is not significant in column (3) of Table A.3. However, this difference is likely a result of the smaller sample we use here: we consider only the 275 SAP user firms among the Fortune 1,000 in this study. Indeed, when we run the same model using the Fortune 1,000 sample, we get a positive and significant estimate of the spillover pool, consistent with Tambe and Hitt (2014b).

References

Aiken LS, West SG, Reno RR (1991) *Multiple Regression: Testing and Interpreting Interactions* (Sage Publications, New York).
Alcácer J, Chung W (2007) Location strategies and knowledge spillovers. *Management Sci.* 53(5):760–776.
Anderson MC, Banker RD, Ravindran S (2003) The new productivity paradox. *Comm. ACM* 46(3):91–94.
Aral S, Weill P (2007) IT assets, organizational capabilities, and firm performance: How resource allocations and organizational differences explain performance variation. *Organ. Sci.* 18(5):763–780.
Aral S, Brynjolfsson E, Wu L (2012) Three-way complementarities: Performance pay, human resource analytics, and information technology. *Management Sci.* 58(5):913–931.
Arellano M, Bond S (1991) Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Rev. Econom. Stud.* 58(2):277–297.

Armstrong CP, Sambamurthy V (1999) Information technology assimilation in firms: The influence of senior leadership and IT infrastructures. *Inform. Systems Res.* 10(4):304–327.

Attewell P (1992) Technology diffusion and organizational learning: The case of business computing. *Organ. Sci.* 3(1):1–19.

Audretsch DB, Feldman MP (1996) Innovative clusters and the industry life cycle. *Rev. Indust. Organ.* 11(2):253–273.

Ba S, Nault BR (2017) Emergent themes in the interface between economics of information systems and management of technology. *Production Oper. Management* 26(4):652–666.

Balli HO, Sørensen BE (2013) Interaction effects in econometrics. *Empirical Econom.* 45(1):583–603.

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

Becker GS (2008) Human capital. The concise encyclopedia of economics. The Library of Economics and Liberty. Accessed December 22, 2021, http://www.econlib.org/library/Enc/HumanCapital.html.

Benkard CL (2000) Learning and forgetting: The dynamics of aircraft production. *Amer. Econom. Rev.* 90(4):1034–1054.

Bloom N, Sadun R, Van Reenen J (2012) Americans do IT better: US multinationals and the productivity miracle. *Amer. Econom. Rev.* 102(1):167–201.

Blundell R, Bond S (1998) Initial conditions and moment restrictions in dynamic panel data models. *J. Econometrics* 87(1):115–143.

Boudreau K, Lakhani K (2009) How to manage outside innovation: Building communities or markets of external innovators? *MIT Sloan Management Rev.* 50(5):69–76.

Bresnahan T, Greenstein S (1996) Technical progress and co-invention in computing and in the uses of computers. *Brookings Papers Econom. Activity Microeconomics* 1996:1–83.

Bresnahan TF, Brynjolfsson E, Hitt LM (2002) Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. *Quart. J. Econom.* 117(1):339–376.

Brynjolfsson E, Hitt L (1995) Information technology as a factor of production: The role of differences among firms. *Econom. Innovation New Tech.* 3(3–4):183–200.

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

Brynjolfsson E, Hitt LM (2003) Computing productivity: Firm-level evidence. *Rev. Econom. Statist.* 85(4):793–808.

Brynjolfsson E, Milgrom P (2012) Complementarity in organizations. Gibbons R, Roberts J, eds. *The Handbook of Organizational Economics* (Princeton University Press, Princeton, NJ), 11–55.

Brynjolfsson E, Fitoussi D, Hitt L (2005) The information technology iceberg. MIT Center for Digital Business Working Paper, Cambridge, MA.

Brynjolfsson E, Hitt LM, Yang S (2002) Intangible assets: Computers and organizational capital. *Brookings Papers Econom. Activity* 2002(1):137–181.

Caselli F, Coleman WJ (2001) Cross-country technology diffusion: The case of computers. *Amer. Econom. Rev.* 91(2):328–335.

Chang YB, Gurbaxani V (2012a) The impact of IT-related spillovers on long-run productivity: An empirical analysis. *Inform. Systems Res.* 23(3):868–886.

Chang YB, Gurbaxani V (2012b) Information technology outsourcing, knowledge transfer, and firm productivity: An empirical analysis. *Management Inform. Systems Quart.* 36(4):1043–1063.

Cheng Z, Nault BR (2007) Industry level supplier-driven IT spillovers. *Management Sci.* 53(8):1199–1216.

Cheng Z, Nault BR (2012) Relative industry concentration and customer-driven IT spillovers. *Inform. Systems Res.* 23(2):340–355.

Chwelos P, Ramirez R, Kraemer KL, Melville NP (2010) Does technological progress alter the nature of information technology as a production input? New evidence and new results. *Inform. Systems Res.* 21(2):392–408.

Cohen WM (2010) Fifty years of empirical studies of innovative activity and performance. Hall BH, Rosenberg N, eds. *Handbook of the Economics of Innovation* (Chapter 4), vol. 1 (Elsevier, Amsterdam), 129–213.

Cohen WM, Levinthal DA (1989) Innovation and learning: The two faces of R&D. *Econom. J. (London)* 99(397):569–596.

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

Cowan R, David PA, Foray D (2000) The explicit economics of knowledge codification and tacitness. *Indust. Corporate Change* 9(2):211–253.

Davenport TH (2000) *Mission Critical: Realizing the Promise of Enterprise Systems* (Harvard Business Press, Boston).

Dewan S, Min C-K (1997) The substitution of information technology for other factors of production: A firm level analysis. *Management Sci.* 43(12):1660–1675.

Dewan S, Shi C, Gurbaxani V (2007) Investigating the risk return relationship of information technology investment: Firm-level empirical analysis. *Management Sci.* 53(12):1829–1842.

Dranove D, Forman C, Goldfarb A, Greenstein S (2014) The trillion dollar conundrum: Complementarities and health information technology. *Amer. Econom. J. Econom. Policy* 6(4):239–270.

Fahey R, Vasconcelos AC, Ellis D (2007) The impact of rewards within communities of practice: A study of the SAP online global community. *Knowledge Management Res. Practice* 5:186–198.

Foerderer J (2020) Interfirm exchange and innovation in platform ecosystems: Evidence from Apple's world wide developers conference. *Management Sci.* 66(10):4772–4787.

Foerderer J, Kude T, Mithas S, Heinzl A (2018) Does platform owner's entry crowd out innovation? Evidence from Google photos. *Inform. Systems Res.* 29(2):444–460.

Gao S, Yeoh W, Wong SF, Scheepers R (2017) A literature analysis of the use of absorptive capacity construct in IS research. *Internat. J. Inform. Management* 37(2):36–42.

Greenstein S, Nagle F (2014) Digital dark matter and the economic contribution of Apache. *Res. Policy* 43(4):623–631.

Griliches Z (1979) Issues in assessing the contribution of research and development to productivity growth. *Bell J. Econom.* 10(1):92–116.

Gu B, Xue L, Ray GN (2008) IT governance and IT investment performance: An empirical analysis. *Proc. 29th Internat. Conf. Inform. Systems, Paris, France*.

Han K, Chang YB, Hahn J (2011) Information technology spillover and productivity: The role of information technology intensity and competition. *J. Management Inform. Systems* 28 (1):115–146.

Hitt LM, Brynjolfsson E (1996) Productivity, business profitability, and consumer surplus: Three different measures of information technology value. *Management Inform. Systems Quart.* 20(2):121–142.

Hitt LM, Wu DJ, Zhou X (2002) Investment in enterprise resource planning: Business impact and productivity measures. *J. Management Inform. Systems* 19(1):71–98.

Howe J (2008) *Crowdsourcing: How the Power of the Crowd Is Driving the Future of Business* (Random House, New York).

Huang P, Zhang Z (2016) Participation in open knowledge communities and job-hopping: Evidence from enterprise software. *Management Inform. Systems Quart.* 40(3):785–806.

Ichniowski C, Shaw K (2003) Beyond incentive pay: Insiders' estimates of the value of complementary human resource management practices. *J. Econom. Perspect.* 17(1):155–180.

Jaccard J, Turrisi R, Wan CK (1990) *Interaction Effects in Multiple Regression (Quantitative Applications in the Social Sciences)* (Sage Publications, New York).

Ko DG, Kirsch LJ, King WR (2005) Antecedents of knowledge transfer from consultants to clients in enterprise system implementations. *Management Inform. Systems Quart.* 29(1):59–85.

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

Lakhani KR, von Hippel E (2003) How open source software works: "Free" user-to-user assistance. *Res. Policy* 32(6):923–943.

Lakhani KR, Wolf R (2005) Why hackers do what they do: Understanding motivation and effort in free/open source software projects. Feller J, Fitzgerald B, Hissam S, Lakhani KR, eds. *Perspectives on Free and Open Source Software* (MIT Press, Cambridge, MA), 3–21.

Lakhani KR, Iansiti M, Fisher N (2014) SAP 2014: Reaching for the cloud. Harvard Business School Case N1-614-052, Cambridge, MA.

Lee DMS, Trauth EM, Farwell D (1995) Critical skills and knowledge requirements of IS professionals: A joint academic/industry investigation. *Management Inform. Systems Quart.* 19(3):313–340.

Levy F, Murnane RJ (2005) *The New Division of Labor: How Computers Are Creating the Next Job Market* (Princeton University Press, Princeton, NJ).

Lieberman MB (1984) The learning curve and pricing in the chemical processing industries. *RAND J. Econom.* 15(2):213–228.

McAfee A (2002) The impact of enterprise information technology adoption on operational performance: An empirical investigation. *Production Oper. Management* 11(1):33–53.

Mehra A, Dewan R, Freimer M (2011) Firms as incubators of open-source software. *Inform. Systems Res.* 22(1):22–38.

Mirchandani V (2014) *SAP Nation: A Runaway Software Economy* (Deal Architect Inc.).

Mortensen PS, Bloch CW (2005) *Oslo Manual—Guidelines for Collecting and Interpreting Innovation Data* (Organisation for Economic Cooporation and Development).

Mun SB, Nadiri MI (2002) Information technology externalities: Empirical evidence from 42 US industries. NBER Working Paper No. 9272, National Bureau of Economic Research, Cambridge, MA.

Nagle F (2018) Learning by contributing: Gaining competitive advantage through contribution to crowdsourced public goods. *Organ. Sci.* 29(4):569–587.

Nagle F (2019) Open source software and firm productivity. *Management Sci.* 65(3):1191–1215.

Pavlou PA, El Sawy OA (2006) From IT leveraging competence to competitive advantage in turbulent environments: The case of new product development. *Inform. Systems Res.* 17(3):198–227.

Roberts N, Galluch PS, Dinger M, Grover V (2012) Absorptive capacity and information systems research: Review, synthesis, and directions for future research. *Management Inform. Systems Quart.* 36(2):625–648.

Roodman D (2009) How to do xtabond2: An introduction to difference and system GMM in Stata. *Stata J.* 9(1):86–136.

Sambamurthy V, Bharadwaj A, Grover V (2003) Shaping agility through digital options: Reconceptualizing the role of information technology in contemporary firms. *Management Inform. Systems Quart.* 27(2):237–263.

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

Tambe P (2014) Big data investment, skills, and firm value. *Management Sci.* 60(6):1452–1469.

Tambe P, Hitt LM (2012) The productivity of information technology investments: New evidence from IT labor data. *Inform. Systems Res*. 23(3-part-1):599–617.

Tambe P, Hitt LM (2014a) Job hopping, information technology spillovers, and productivity growth. *Management Sci.* 60(2):338–355.

Tambe P, Hitt LM (2014b) Measuring information technology spillovers. *Inform. Systems Res.* 25(1):53–71.

Tambe P, Hitt LM, Brynjolfsson E (2012) The extroverted firm: How external information practices affect innovation and productivity. *Management Sci.* 58(5):843–859.

Teece DJ (1977) Technology transfer by multinational firms: The resource cost of transferring technological know-how. *Econom. J. (London).* 87(346):242–261.

Thornton RA, Thompson P (2001) Learning from experience and learning from others: An exploration of learning and spillovers in wartime shipbuilding. *Amer. Econom. Rev.* 91(5):1350–1368.

Tilton JE (1971) *International Diffusion of Technology: The Case of Semiconductors* (Brookings Institution Press, Washington D.C.).

Trantopoulos K, von Krogh G, Wallin MW, Woerter M (2017) External knowledge and information technology: Implications for process innovation performance. *Management Inform. Systems Quart.* 41(1):287–300.

von Hippel E (1994) Sticky information and the locus of problem solving: Implications for innovation. *Management Sci.* 40(4): 429–439.

von Hippel E, Tyre MJ (1995) How learning by doing is done: Problem identification in novel process equipment. *Res. Policy* 24(1):1–12.

Walker BASBB (2012) *Successfully Navigating the Turbulent Skies of a Large-Scale ERP Implementation* (HEC Montreal Centre for Case Studies).

Wu L, Hitt LM, Lou B (2020) Data analytics skills, innovation and firm productivity. *Management Sci.* 66(5):2017–2039.

Wu L, Jin F, Hitt LM (2017) Are all spillovers created equal? A network perspective on information technology labor movements. *Management Sci.* 64(7):3168–3186.

Xu Q, Ma Q (2008) Determinants of ERP implementation knowledge transfer. *Inform. Management* 45(8):528–539.

Yusuf Y, Gunasekaran A, Abthorpe MS (2004) Enterprise information systems project implementation: A case study of ERP in Rolls-Royce. *Internat. J. Production Econom.* 87(3):251–266.