

# The Illusory Diffusion of Innovation: An Examination of Assimilation Gaps

Robert G. Fichman • Chris F. Kemerer

Wallace E. Carroll School of Management, Boston College, 354D Fulton Hall, 140 Commonwealth Avenue,  
Chestnut Hill, Massachusetts 02167-3808

fichman@bc.edu

University of Pittsburgh, 278a Mervis Hall, Pittsburgh, Pennsylvania 15260

ckemerer@mail.business.pitt.edu

---

Innovation researchers have known for some time that a new information technology may be widely acquired, but then only sparsely deployed among acquiring firms. When this happens, the observed pattern of cumulative adoptions will vary depending on which event in the assimilation process (i.e., acquisition or deployment) is treated as the adoption event. Instead of mirroring one another, a widening gap—termed here an *assimilation gap*—will exist between the cumulative adoption curves associated with the alternatively conceived adoption events. When a pronounced assimilation gap exists, the common practice of using cumulative purchases or acquisitions as the basis for diffusion modeling can present an illusory picture of the diffusion process—leading to potentially erroneous judgments about the robustness of the diffusion process already observed, and of the technology's future prospects. Researchers may draw inappropriate theoretical inferences about the forces driving diffusion. Practitioners may commit to a technology based on a belief that pervasive adoption is inevitable, when it is not.

This study introduces the assimilation gap concept, and develops a general operational measure derived from the difference between the cumulative acquisition and deployment patterns. It describes how two characteristics—*increasing returns* to adoption and *knowledge barriers* impeding adoption—separately and in combination may serve to predispose a technology to exhibit a pronounced gap. It develops techniques for measuring assimilation gaps, for establishing whether two gaps are significantly different from each other, and for establishing whether a particular gap is absolutely large enough to be of substantive interest. Finally, it demonstrates these techniques in an analysis of adoption data for three prominent innovations in software process technology—relational database management systems (RDBs), general purpose fourth generation languages (4GLs), and computer aided software engineering tools (CASE). The analysis confirmed that assimilation gaps can be sensibly measured, and that their measured size is largely consistent with a priori expectations and recent research results. A very pronounced gap was found for CASE, while more moderate—though still significant—gaps were found for RDBs and 4GLs.

These results have the immediate implication that, where the possibility of a substantial assimilation gap exists, the time of deployment should be captured instead of, or in addition to, time of acquisition as the basis for diffusion modeling. More generally, the results suggest that observers be guarded about concluding, based on sales data, that an innovation is destined to become widely used. In addition, by providing the ability to analyze and compare assimilation gaps, this study provides an analytic foundation for future research on why assimilation gaps occur, and what might be done to reduce them.

(*Assimilation Gap; Software Process Innovation; Adoption; Deployment; Diffusion Modeling*)

---

## 1. Introduction

For innovations in information technology to have a positive impact on quality and productivity, they must actually be deployed. Yet, innovation researchers have known for some time that a new technology may be introduced amid great enthusiasm and enjoy widespread initial acquisition, but nevertheless still fail to be thoroughly deployed among many acquiring firms. Liker et al. report that Computer Aided Design (CAD) technologies had achieved unusually rapid market penetration in the 1980s, yet, as late as 1992 "true CAD/CAM [utilization was] still quite rare" (Liker et al. 1992, p. 75) Cooper and Zmud, in a study of material requirements planning (MRP), report that, while 73% of surveyed companies were using MRP, only 27% of respondents had progressed beyond Class C MRP implementation, a relatively low level of utilization (Cooper and Zmud 1990). Eveland and Tornatzky, in describing the fate of machine vision systems (a popular innovation introduced in the late 1970s) observe that: "Many plants simply gave up. Some large, and expensive machine vision systems were 'de-installed.' Automation consultants, in visits to plants, found unused machine vision systems sitting in boxes, relics of failed deployment" (Eveland and Tornatzky 1990, p. 123).

These examples illustrate the following basic insight: widespread acquisition of an innovation need not be followed by widespread deployment and use by acquiring organizations. While the implications of this insight have been incorporated into some of the more recent studies focusing on the antecedents of organizational innovation (Meyer and Goes 1988, Cooper and Zmud 1990), there has not been a comparable recognition by researchers engaged in modeling macro-level patterns of diffusion, even though the implications for diffusion modeling are arguably equally important.

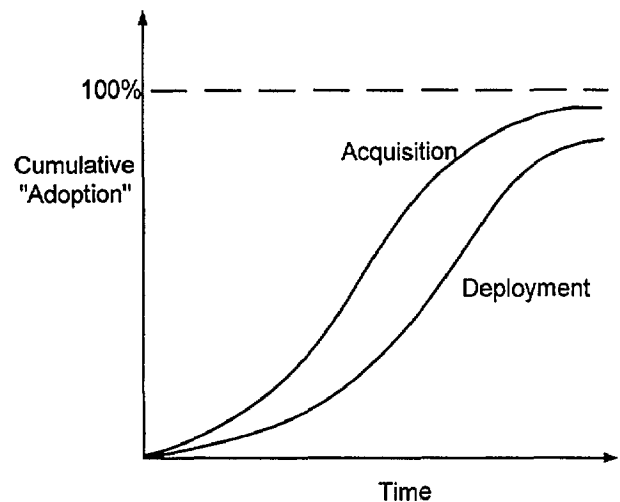
Diffusion modeling studies are concerned with understanding the patterns innovations follow as they spread across a population of potential adopters over time. A typical approach is to define adoption as the purchase or physical acquisition of the innovation, and then to fit a times series of observed cumulative adoption counts or percentages to some functional form, such as the logistic distribution (Mahajan and Peterson

1985). Some studies seek to infer support for alternative theories of diffusion based on the observed pattern of adoption for a particular innovation (Brancheau and Wetherbe 1990, Gurbaxani 1990, Gurbaxani and Mendelson 1990). Others compare multiple innovations, seeking to explain why some innovations diffuse more rapidly and widely than others (Mansfield 1993). Still others have a more applied focus and seek to make predictions about the future course of innovation for a technology (Mahajan et al. 1990).

When it can be safely assumed that later events in the process of intraorganizational assimilation will nearly always follow quickly on the heels of earlier events, then the observed pattern of cumulative adoptions will not vary much depending on the particular assimilation event used to define the time of adoption. The diffusion pattern that results when an earlier event (e.g., acquisition) is used will closely mirror the pattern that results when a later event is used (e.g., deployment), as illustrated in Figure 1. In this case, the conclusions a particular study draws are not likely to be contingent on the definition employed for adoption.

However, for some technologies it may be inappropriate to assume that in most organizations these later assimilation events will automatically follow earlier events. As a result, the pattern of cumulative deployments may not closely mirror the pattern of cumulative acquisitions, but rather, there may be a widening "gap" between the two curves plotted as a function of

Figure 1 Diffusion Curves for Alternative Definitions of "Adoption"



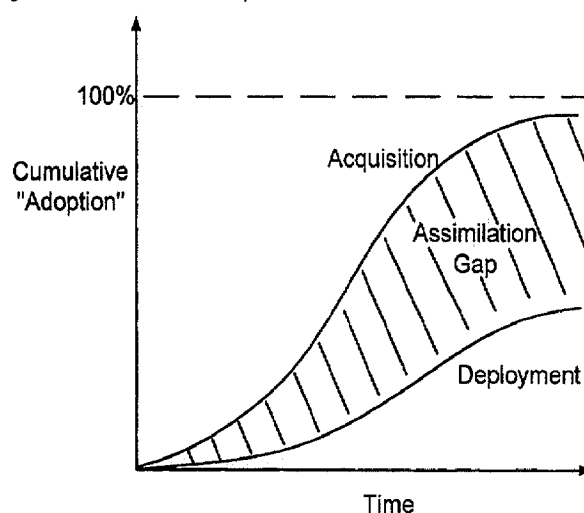
time (see Figure 2). Because this gap is bounded by the cumulative adoption curves associated with two alternative assimilation events, we label it the *assimilation gap*.

When a substantial assimilation gap exists for an innovation, the use of cumulative acquisition as the basis for diffusion modeling can present an illusory picture of the diffusion process—leading to potentially erroneous judgments about the robustness of the diffusion process already observed, and about the technology's future prospects. Researchers may draw inappropriate theoretical inferences about the forces driving diffusion. Practitioners may commit to a technology based on a belief that pervasive adoption is inevitable, when it is not.

Prior innovation research lays the groundwork for several potential explanations for why some innovations might be prone to an especially large assimilation gap. For example, it could be that high knowledge barriers, which have been found to generally slow diffusion, tend to have an especially negative effect on deployment compared with acquisition (Attewell 1992). Alternatively, it could be that the potential option value to use some innovations in the future is so high that many organizations are willing to initiate deployment simply to preserve this option (Cohen and Levinthal 1990). Or, perhaps something about the way some innovations are marketed leads many organizations to acquire technologies under one set of expectations, only to subsequently encounter a much different, less favorable reality (Rosenberg 1976). Whatever the reason for the existence of pronounced assimilation gaps, the first step toward using the concept to make predictions or to test rival theories is to develop a foundation of definitions, measures, and analytical techniques for the concept. The purpose of this research is to provide such a foundation. With this in place, researchers will be in a position to investigate a number of important questions, such as:

- (1) Is an observed assimilation gap large enough to be of substantive interest?
- (2) Is the observed assimilation gap for one technology significantly larger than for another?
- (3) Why does a particular technology have a substantial assimilation gap?

Figure 2 Assimilation Gap



(4) Why do two technologies have significantly different gaps?

(5) Why do two adopter populations for the same technology have significantly different gaps?

In this article we address the first two questions, and suggest alternative avenues for investigating the other three. We begin by formalizing the assimilation gap concept and proposing an operational definition (§2). We then present several explanations for why some technologies—including those in this study—might be especially prone to assimilation gaps (§3). Next, we provide an empirical examination, based on both graphical and survival analysis techniques, of the assimilation gap for three prominent IT innovations: relational database management systems (RDBs), fourth generation languages (4GLs), and computer aided software engineering tools (CASE) (§4–§5). Finally, we discuss these empirical results and draw conclusions for researchers and practitioners (§6–§7).

## 2. The Assimilation Gap Concept

When later events in the assimilation process do not follow quickly or reliably from earlier events, the result is an assimilation gap. For this study we propose two different, but complementary ways to conceive the assimilation gap. Our primary approach is based on traditional diffusion modeling concepts, and has a

straightforward interpretation as the difference between two cumulative adoption curves for the population as a whole. Our second approach is based on survival analysis concepts, and considers the durations of the lags between acquisition and deployment within acquiring firms. The first approach maps most directly to our theoretical construct, while the second lays a foundation for key statistical inferences, such as that the magnitudes of two assimilation gaps significantly differ.

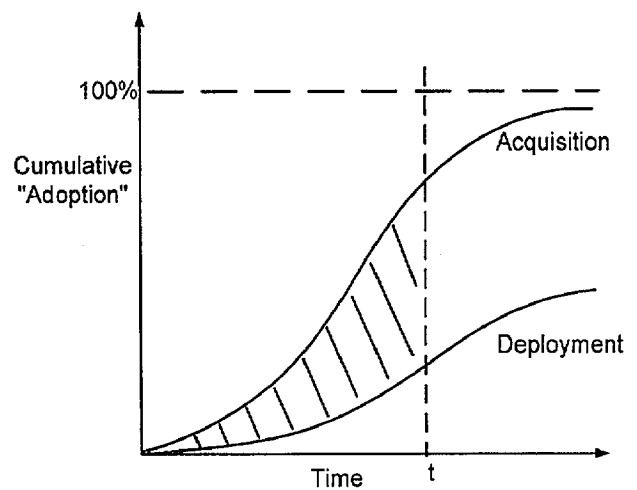
### 2.1. The Diffusion Modeling Concept of the Assimilation Gap

Diffusion is a process whereby an innovation spreads across a population of potential adopters over time (Rogers 1995). The process begins with introduction of the innovation to the population, and ends when the population becomes saturated with adoptions, i.e., when all those who will ever adopt have, in fact, adopted. Saturation may occur at close to 100% of the population, or may fall far short of this. The primary tool for analysis of diffusion patterns is the cumulative adoption curve, which shows the percentage of a population that has adopted at any given point during the diffusion process.

The assimilation gap occurs when the pattern of adoptions based on one assimilation event is at odds with the pattern for a later assimilation event. For this study we define the assimilation gap as *the difference between the pattern of cumulative acquisitions and cumulative deployments of an innovation across a population of potential adopters*. Although this definition is made in reference to two particular events—acquisition and deployment—in principle, any two assimilation events could be used to define an assimilation gap, if so warranted by the objectives of the research.

For an operational measure of the assimilation gap we propose *the area between the cumulative acquisition and cumulative deployment curves at time  $T$  as a proportion of the area under the cumulative acquisition curve at time  $T$* . As illustrated in Figure 3, this measure is computed as the cross-hatched area divided by the area under the cumulative acquisition curve. Although a large assimilation gap (as we operationalize it here) may sometimes occur even when acquisition of a technology has been slow—as long as deployment is much slower

Figure 3 Computing the Assimilation Gap



still—we focus on the case where a large assimilation gap occurs for a technology that has been rapidly acquired. This is so for two reasons. First, from a practical standpoint, it means that a larger proportion of the population may be affected by the gap. Second, we believe that this case raises the most interesting theoretical questions, since the technology appears “successful” according to one view of its diffusion (based on acquisition) but much less so according to another (based on deployment). In this situation, the illusory quality of diffusion is at its peak.

### 2.2. The Survival Analysis Concept of the Assimilation Gap

With the survival analysis concept of the assimilation gap, we depart from traditional diffusion modeling techniques and look at the deployment process within acquiring firms. Specifically, we consider the time it takes for firms that have already acquired an innovation to actually deploy it. The longer the typical acquiring firm takes to deploy an innovation once acquired, the larger the assimilation gap will be.

Survival analysis techniques were originally developed by biostatisticians studying the duration of human lifetimes, hence the label, “survival analysis,” although, more recently, these techniques have received growing attention by social scientists for use in studies that share the central problem presented by lifetime data, namely, the censoring of event times (Lawless 1982, Cox and Oakes 1984, Singer and Willett 1991).

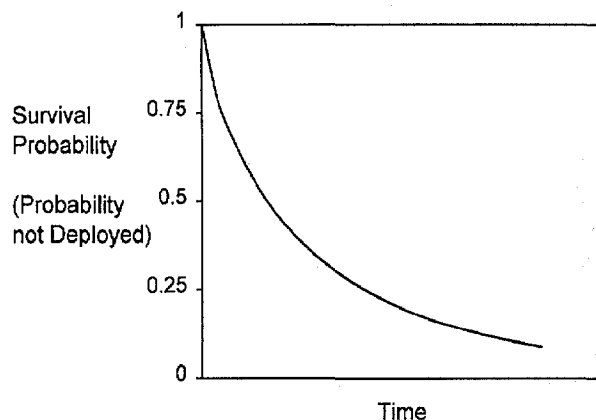
Censoring occurs when some units in a study have not exhibited the event of interest during the study's observation period, and so all that is known for those units is that the event times will exceed the durations for which the units have been observed.<sup>1</sup> In the current study censoring has occurred whenever a firm has acquired but not yet deployed during the observation window.

One of the key tools in survival analysis—and the one of particular interest here—is the survivor function (see Figure 4). The survivor function provides a summary view of the event times for a population. More specifically, the survivor function provides an estimate of the proportion of a population expected to have an event time exceeding any given time  $T$ . In this case we are interested in the time it takes for an acquiring firm to deploy a technology, so the "survivor function" can be thought of as the "survival" of the earlier technology, despite acquisition of the new technology.

In more typical applications—such as those involving the time until death, job turnover, or component failure—long event times, as evidenced by a slowly decreasing survivor function, are assumed to be desirable. In our current application, the opposite is true: long event times (i.e., the time it takes acquiring firms to deploy) are assumed to be undesirable, since this

<sup>1</sup>With traditional statistical techniques, such as regression analysis, censoring of data leads to biases in the estimation of parameters; sensitivity to the length of the data collection period is also common. Survival analysis, by contrast, makes appropriate use of all observations, even censored ones.

Figure 4 The Survivor Function



represents the time it takes for firms to get significant benefits from a technology they have purchased.

The survivor analysis concept of the assimilation gap, as embodied in the survivor function for deployment times, relates to our graphically-based, diffusion modeling concept as follows. A rapidly decreasing survivor function corresponds to a small assimilation gap, while a slowly decreasing survivor curve corresponds to a large gap. The benefits of introducing survivor analysis on top of diffusion modeling are two fold. First, it opens up new kinds of analyses, such as estimating the expected median time to deployment, and the expected fraction of acquirers that will never deploy. Second, and more importantly, it provides a means to draw statistical inferences on the relative sizes of two observed assimilation gaps.

### 3. Theoretical Explanations for the Assimilation Gap

As mentioned earlier, prior work in the innovation field lays the foundation for several theoretical explanations accounting for the propensity of certain innovations to exhibit a pronounced assimilation gap. Since an assimilation gap is a phenomenon that occurs at the level of a technology, we look primarily to distinctive characteristics of the technology itself—and of the institutional environmental supporting it—for plausible explanations. This is in keeping with other prominent theories about what drives the overall pattern of innovation diffusion (Tornatzky and Fleischer 1990, Attewell 1992, Rogers 1995).

We have previously argued that technologies used in the software development process—what we refer to as *software process innovations* (SPIs)<sup>2</sup>—possess two characteristics that predispose them to significant assimilation gaps: strongly *increasing returns* to adoption, and substantial *knowledge barriers* impeding adoption (Fichman and Kemerer 1993a). These arguments, summarized below, lead us to choose prominent SPIs for the empirical portion of this study. This is not to imply

<sup>2</sup>We refer to these technologies as software process innovations because, when acquired and deployed, they change an organization's process for developing software applications. In Swanson's typology of information systems innovations, an SPI would be classified as a Type Ia or Ib "IS process" innovation (1994).

that other factors (e.g., structural, managerial, political, social) have no effect on the acquisition and deployment of IT innovations within particular firms—they certainly do (Kwon and Zmud 1987, Fichman 1992, Prescott 1995, Prescott and Conger 1995). However it must be remembered that we are not attempting to explain why a particular firm has experienced rapid acquisition followed by slow deployment, but rather, why a population of firms has exhibited this pattern, and more to the point, why there are substantial differences in this pattern across technologies. It is possible that other factors beyond increasing returns and knowledge barriers might (a) be systematically present among some population of potential adopters, (b) be more salient for some kinds of innovations than for others, and (c) have a strongly differential effect on acquisition versus deployment. In this case, an assimilation gap should also result. In the discussion section, we touch on some other candidate factors that appear less salient in the case of SPIs, but may well apply in other situations.

### 3.1. Increasing Returns to Adoption

Some technologies become much more valuable to a given adopter to the extent that others also adopt. Such technologies are said to be subject to increasing returns to adoption (Arthur 1988, Arthur 1996). Increasing returns arise from the incremental contribution additional adopters make to the accumulated benefits of: (1) positive network externalities among adopters (Katz and Shapiro 1986, Markus 1987), (2) learning-by-using among adopters (Rosenberg 1982), (3) economies of scale in production and learning-by-doing among producers (Arrow 1962), (4) general industry knowledge about the innovation (Arthur 1988), and (5) a more rapidly maturing technology infrastructure (Arthur 1988, Van de Ven 1993).<sup>3</sup> Many information technologies are subject to increasing returns (Brynjolfsson and Kemerer 1997, Schilling 1998) and this is particularly true of the innovations used in the software development process. In the case of RDBs, for example, a large adoption base means that suppliers will have a greater base of customers over which to spread R&D costs, thus accelerating improvement of

the technology. A large base of RDB users can also be expected to attract the attention of third-party developers, including those that specialize in integrating databases with other technologies (such as languages and tools). More widespread adoption also implies more and larger RDB users groups to share knowledge about the technology. Finally, a large base of RDB adopters means greater availability of experienced staff and training and consulting services.<sup>4</sup>

The presence of substantial increasing returns means that, by definition, a wide discrepancy will exist between a technology's *initial performance* (which can be thought of as the performance an average adopter is likely to achieve with a technology during its first few years of commercial availability, including not only productivity or quality improvements or decrements, but also amortized learning and disruption costs), and its *network potential* (which can be thought of as the hypothetical future performance that could be achieved if a technology were to become universally adopted by the network of users, suppliers, and mediating institutions). This discrepancy lays the foundation for two theoretical explanations for the assimilation gap, one normatively rational, and one less so.

The normatively rational argument follows from the work of Cohen and Levinthal on the role of absorptive capacity in R&D and innovation (Cohen and Levinthal 1990). They argue that absorptive capacity—the ability to assimilate an innovation and apply it to productive ends—is cumulative and path dependent. Because organizations cannot instantaneously assimilate an innovation at the exact point that its benefits have become certain and obvious, they may choose to begin the assimilation process even when the ultimate benefits of using the technology are unclear. This explanation suggests that the assimilation gap flows from conscious attempts by managers to hedge their technological bets in the face of a dynamic and uncertain technology. In other words, managerial decision making is being driven by a real options perspective on adoption (McGraith 1997).

The less normatively rational argument is based on

<sup>3</sup>Arthur (1988) provides an extended discussion of each of these sources of increasing returns.

<sup>4</sup>See Fichman and Kemerer (1993a), (1993b) for more detailed development of the theoretical linkage between SPIs and increasing returns.

the concept of "signaling" (Attewell 1992). In the case of some commercial technologies, vendors and other stakeholders may communicate a vision of an emerging technology's main features and benefits based on what it would be like to use the technology at its network potential, rather than its likely initial performance level. Claims about productivity or quality improvements may fail to incorporate the adverse impact of technological immaturity, the absence of complementary tools, and the difficulty of hiring and training—and thus fail to communicate to potential adopters that these can be pertinent considerations. Organizations may decide to initiate the assimilation process based on the positive vision of what the technology could become, only to subsequently learn that the current state of the technology is not as mature or robust as claimed. One of the more frequent comments we encountered in our recent study on the adoption of object technology was surprise and displeasure at the inadequacies of supporting processes and tools (Fichman and Kemerer 1997b). Similar unpleasant surprises have also been reported for CASE tools (McComb 1994, Martin 1995, Senn and Wynekoop 1995). Unexpected immaturity and related problems can discourage deployment directly by impeding the implementation process, or indirectly by persuading adopters to scale back or terminate efforts to implement what they have come to believe is an "over sold" technology (Senn and Wynekoop 1995).

### 3.2. Knowledge Barriers

For some kinds of technologies, adoption and use is hindered by the effort of organizational learning required to obtain necessary knowledge and skills. When this occurs, the technology is said to be subject to knowledge barriers (Attewell 1992). Knowledge barriers arise because the technological and managerial knowledge required to successfully deploy complex technologies typically goes far beyond simple awareness of the innovation and its potential benefits. Such knowledge tends to be "sticky" (von Hippel 1994), and is usually acquired only over time, and with considerable difficulty (Cohen and Levinthal 1990, Kogut and Zander 1992). As a class, SPIs certainly qualify as the sort of "complex organizational technologies" Attewell (1992) had in mind when developing his

knowledge barrier theory of diffusion. According to recent reports, this is especially true of CASE (Rai 1995, Martin 1995, Senn and Wynekoop 1995). The assimilation of CASE tools requires organizational learning on a number of separate fronts, including: learning about the complex underlying methodologies automated by the CASE tools (Fichman and Kemerer 1993b); learning about the CASE philosophy, which advocates an enterprise model-based approach to systems development (Stone 1993); learning how to use the actual software products that provide CASE functionality (Kemerer 1992); and finally, learning how to structure project teams and incentives in light of the potentially radically different approaches to development that often accompany CASE (Orlikowski 1993).<sup>5</sup>

When knowledge barriers are present, many of the organizations that choose to acquire the innovation may be unable or unwilling to deploy it. This is because the organizational knowledge needed to generalize, scale up, and institutionalize a technology differs not only in magnitude but in kind from the knowledge needed to acquire it (Leonard-Barton 1989). As was the case with increasing returns, the existence of knowledge barriers may create a kind of discrepancy that ultimately leads to an assimilation gap. In this case, the discrepancy is the differential (i.e., larger) impact of knowledge barriers on deployment versus acquisition.

A reasonable question one might ask at this point is why most organizations would not anticipate the difficulties flowing from knowledge barriers, and, as Attewell (1992) has found often occurs, simply defer acquisition until knowledge barriers have been sufficiently lowered. There are two possible explanations. First, organizations may choose to knowingly take on risky ventures, as argued above, because this may be necessary to create the option of being in a position to immediately use the technology when the appropriate time has arrived. Second, the inherent nature of knowledge barriers themselves precludes the perfect foresight needed to anticipate the magnitude of the challenges associated with deployment. The fact that the organization is not knowledgeable about the innovation in general means they are also lacking in specific

<sup>5</sup>See Fichman and Kemerer (1993a, 1997a, 1997b) for a more detailed development of the linkage between SPIs and knowledge barriers.

knowledge about how to best implement it. Either way, the assimilation gap ultimately results from the differentially adverse impact of knowledge barriers on deployment versus acquisition.

## 4. Modeling Assimilation Gaps for Software Process Innovations

We now present an examination of assimilation gaps for three of the more prominent information technologies to emerge in the 1980s: relational database management systems (RDBs), general purpose fourth generation languages (4GLs), and Computer Aided Software Engineering tools—tools that support systems analysis and design (CASE). These three software process innovations were selected to support the analysis of assimilation gaps because of our a priori expectation that they would exhibit significant assimilation gaps (Fichman and Kemerer 1993b), and because of the expectations that there would be contrasts among them in the size of measured gaps. In particular, we expected CASE would have an especially large gap, based on reports at the time of implementation difficulties and under-utilization (Howard and Rai 1993).

### 4.1. Study Methods

Data were gathered through a large scale cross-sectional survey of 1,500 medium to large organizations, with over 500 total employees (Fichman 1995). The sampling unit was the IT department at individual sites. Informants were instructed to consider just their own site in answering questions. A probability sample of 1,500 sites was extracted from a list, maintained by International Data Corporation, of 40,000 US sites with computers installed.<sup>6</sup> The ideal informants were set out in advance to be middle-level IT managers with a good knowledge of applications development activities and

<sup>6</sup>The goal of this list is to provide advertisers in *ComputerWorld* magazine—the flagship trade publication for the IT industry—with good prospects for their sales efforts. The target informants are IT professionals with buying authority for hardware, software or telecommunications products. All results reported here are based on unweighted data. Case weighting (to counter the effects of probability sampling) was found to have a negligible effect on initial descriptive results, and hence, was viewed as an unnecessary complication for the full analysis.

technologies at their respective sites. Actual respondents overwhelmingly met these criteria.<sup>7</sup> To qualify for the sampling frame, the site had to: 1) be part of a larger enterprise with at least 500 employees; 2) have previously reported at least some tools for software development installed at the site; 3) have an informant with an IT management-related title; and 4) have at least some Microsoft DOS or Windows-based computers on site. The first two criteria increase the likelihood of custom applications development at the site. The third criterion helps ensure a well-informed respondent. The last criterion—that only excluded an incremental 5% of the sample—increases the likelihood that respondents have the means to take the survey (a DOS based computer).

The survey was administered via computer disk, a novel approach whereby respondents insert the survey software into their PCs and are automatically led through the questionnaire items (Saltzman 1993). Numerous strategies, including most of those recommended by Dillman (1983), were employed to boost response rates and eliminate bias.<sup>8</sup> A total of 679 surveys were returned, for a raw response rate of 45%.

To help ensure that informants were operating from a common understanding of what was meant by the terms “relational database management systems,” “fourth-generation language,” and “computer-aided software engineering,” the questionnaire provided a list of several prominent, commercially available instances of each SPI. To promote further consistency in the minds of respondents, the questionnaire narrowed

<sup>7</sup>Nine of ten reported holding IT “manager” or “director” level titles. Three-fourths reported responsibility for at least half of applications development activities at the site. Eighty-four percent reported being “very” or “extremely” knowledgeable about applications development at the site.

<sup>8</sup>These included use of an individually addressed and signed cover letter assuring confidentiality and emphasizing the survey’s university affiliation; provision of incentives, including a chance to receive one of five \$100 lottery prizes, an executive summary of the results and some related technical reports; use of a reminder card and follow-up letter; use of other techniques to communicate our commitment and investment in the survey, including provision of an 800 number for questions and a postage paid, pre-addressed mailer for returning the disks.



the scope to SPIs appropriate for use on medium to large multiuser applications.<sup>9</sup>

An analysis of the descriptive characteristics of responding sites was performed based on 608 usable responses out of a total of 679 returned disks.<sup>10</sup> The vast majority of responding sites were typical corporate information systems organizations, with mainframe or midrange computers as their primary host environment (81%). Smaller departments (16% reported fewer than 5 total staff members) and larger departments (13% reported over 100 IT staff members) were included. The median reported size of the total on-site IT staff (including development, technical support and operations) was 16. A wide range of business sectors was represented, including public and quasi-public sectors (i.e., government, education, and health care).

#### 4.2. Cumulative Adoption Calculations

In modeling the cumulative adoption of an innovation for a population over time, three measurement-related decisions are required: (1) how to define the population; (2) how to define the time of the adoption event for each member of the population; and (3) how to define the total time period to be covered. The population for this study was defined as those respondents to the disk-based survey who report more than five full-time application developers on staff. Since the organizations were randomly selected from a high-quality list, they are believed to be broadly representative of U.S. enterprises. The requirement of more than five developers is intended to confine analysis to organizations that have the application development scale needed to be *prima facie* strong candidates for

adoption of the technologies under study (Fichman and Kemerer 1997a).<sup>11</sup> Eliminating cases with five or fewer developers reduced the sample size from 608 to 395; elimination of 11 additional unusable cases further reduced sample size to 384.<sup>12</sup> Among the retained 384 cases, the levels of acquisition for RDBs, 4GLs, and CASE were 83%, 63%, and 42% respectively, as of the end of 1993. (In the full sample, these percentages were 68%, 51%, and 29%, respectively.)

Two alternative definitions of the adoption event are employed: acquisition and deployment. The time of acquisition is defined as the reported year that an actual instance of the technology was first purchased by the site. The time of deployment is defined as the year that use of the technology first reached 25% of all new development. The first measure is intended to parallel a definition of adoption frequently used in traditional diffusion modeling studies, including all of those based on first-purchase data (see, for example, Bass 1969). The second measure is intended to capture whether the site has gotten beyond a limited level of deployment.<sup>13</sup> The resulting time series for the three technologies are the cumulative number of acquisitions and deployments by the end of each year.

The time periods to be covered by the analysis were defined in order to achieve a consistent base of comparison for the three technologies. The end year was

<sup>9</sup>For example, prior to the block of questions related to CASE tools, respondents were instructed as follows: "In the questions that follow, we are interested only in traditional (nonobject oriented) CASE tools appropriate for developing medium to large multiuser applications, e.g., Texas Instruments IEF, KnowledgeWare ADW, Excelerator, CADRE TeamWork, LDBMS."

<sup>10</sup>Of the 71 unusable responses, 20 were received after the agreed upon cut-off employed by a subcontractor assisting with data collection, 39 did not meet the original criteria for inclusion (either because no applications development was performed at the site or because the manager reported being ill-informed about software development at the site), 7 contained inconsistent responses to key questions (suggesting misunderstanding of those questions), and 5 were incomplete.

<sup>11</sup>The main reasons for confining analysis in this way are to improve the face validity of descriptive results (e.g., reported levels of acquisition), and to counter a concern that statistical conclusions might be affected by the inclusion of a large number of potentially inappropriate adopters. As a check, analysis of data including firms with five or fewer developers does not change the results reported.

<sup>12</sup>Six cases reported deployment prior to acquisition. Five cases reported acquisition in years prior to the commercial availability of the technology.

<sup>13</sup>The 25% level of use is considered to be a reasonable cut-off for these SPIs, as near 100% deployment can be seen as the goal state for all three SPIs. RDBs were promoted as replacement for prior database models, and there are few business applications for which, in principle, databases are not the most appropriate data management tool. Production 4GLs were promoted as a replacement for Cobol, a language that had accounted for 90% of the installed base for business applications when 4GLs were introduced. CASE, which was designed to support a philosophy of model-based development, should, once adopted, be used on every new development project, regardless of implementation environment, lest an incomplete model result.

defined to be 1993 for all three technologies, as this was the last full year for which data were available. To establish the start year for each series, a two-step procedure was followed. First, the year in which cumulative acquisitions first reached at least 10% of the sample was calculated for each, and assigned to be the "anchor year" for that technology. The start year for each series was then assigned to be three years prior to the anchor year. This resulted in start years of 1982, 1981, and 1986 for RDBs, 4GLs, and CASE, respectively for this data set.<sup>14</sup> The resulting cumulative acquisition and deployment series are provided in Table 1. Cumulative adoptions are reported as percentages of the sample population. As Table 1 shows, there are 12 observation periods for RDBs (1982–1993), 13 periods for 4GLs (1981–1993), and 8 periods for CASE (1986–1993).

## 5. Results

This section presents the results of several interrelated analyses. It begins with a descriptive examination of assimilation gaps based on visual inspections of the cumulative acquisition and deployment curves for the three technologies. This is followed by presentation of a quantitative approach to computing and presenting assimilation gap estimates. Finally, survival analysis techniques are used to determine whether the assimilation gaps associated with the technologies are significantly different, and whether each gap—taken by itself—is absolutely large enough to be of substantive interest.

### 5.1. Descriptive Results

The cumulative acquisition and deployment series for RDBs, 4GLs, and CASE are plotted in Figures 5 through 7. To facilitate visual comparisons, the same

<sup>14</sup>The three year rule was designed to include enough prior time periods so that the distinctive "take-off" pattern common to innovation diffusion would be evident, but not so many time periods that chance variation in reported acquisitions would be an overwhelming concern. This represents a more useful method for anchoring a series, than, for example, taking the first year in which any acquisitions occurred. This is because the start year under such a rule could be strongly subject to outliers and/or chance variation. However, it should be pointed out that using a different number of years prior to the anchor year, two or four, had a negligible effect on results.

scales have been used for all three figures, and a vertical line is provided to highlight the eighth year of each series. RDBs appear to have a moderate assimilation gap. By 1993, about 83% ( $N = 317$ ) of sites had acquired an RDB, although only about 42% ( $N = 161$ ) had yet deployed RDBs on at least 25% of new development projects. But, both the acquisition and deployment curves show robust increasing trends, and there is no reason to conclude, based on an inspection of these series, that deployment will not eventually be achieved by a large majority of acquirers.

In visually comparing Figure 5 to Figure 6, production 4GLs appear to have a slightly larger assimilation gap than RDBs.<sup>15</sup> By 1993, about 63% ( $N = 241$ ) had acquired 4GLs, but only 26% ( $N = 99$ ) had deployed. One must go back 6.5 years, to mid-1986, to find the year where cumulative acquisitions equals the number of cumulative deployments in 1993. For RDBs, by comparison, one must go back only 4 years, to 1989, to find the year where cumulative acquisitions equals cumulative deployments in 1993.

For CASE, however, the apparent assimilation gap is striking. By 1993, 42% ( $N = 160$ ) have acquired a CASE tool, but only 7% ( $N = 26$ ) have deployed. Of course, one possible explanation is that this larger apparent gap is an artifact of the shorter observation period. To examine this concern, the acquirer/deployer ratio as of the end of 8 years was computed for all three technologies. This ratio was 6.2 to 1 for CASE, but only 2.2 to 1 and 2.5 to 1, respectively, for RDBs and 4GLs. Therefore, the difference is not believed to result from the shorter observation period for CASE.

To summarize, the descriptive results suggest moderate assimilation gaps for RDBs and 4GLs, and a pronounced gap for CASE.

### 5.2. Quantitative Estimation of Assimilation Gaps

The descriptive analysis presented above provides an intuitive feel for the magnitude of assimilation gaps, and the results appear sensible when compared with prior expectations and recent research results. In this section we present a quantitative analysis of assimilation gaps.

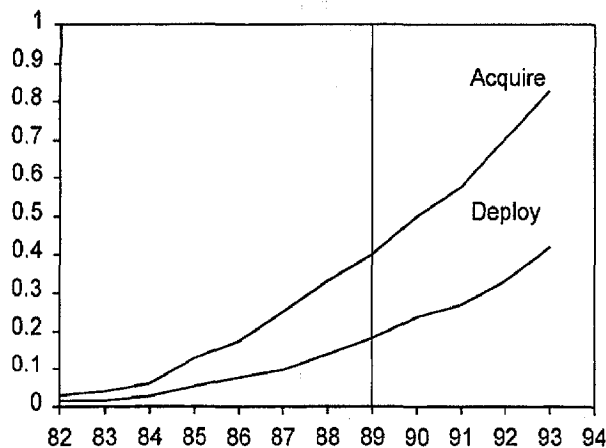
As described previously, the proposed definition of

<sup>15</sup>A statistical comparison provided later in this section, and summarized in Table 3, confirms this difference.

**Table 1** Cumulative Acquisition and Deployment Percents for Three Technologies (*N* = 384)

Calendar Year	RDBs			4GLs			CASE		
	Period	Acquisition	Deployment	Period	Acquisition	Deployment	Period	Acquisition	Deployment
1981				1	3.1	0.8			
1982	1	2.9	1.6	2	4.7	1.0			
1983	2	3.1	1.6	3	8.1	1.3			
1984	3	6.0	2.6	4	10.4	1.8			
1985	4	12.8	5.2	5	19.0	5.5			
1986	5	17.4	7.6	6	22.7	7.0	1	1.0	0.3
1987	6	25.0	9.6	7	28.4	9.6	2	3.1	0.8
1988	7	33.1	14.0	8	31.8	12.5	3	5.5	1.0
1989	8	40.0	18.2	9	37.2	14.3	4	9.9	1.0
1990	9	50.0	23.7	10	42.7	17.2	5	16.7	1.5
1991	10	57.6	26.8	11	47.1	19.2	6	25.0	3.1
1992	11	70.3	33.1	12	55.7	21.6	7	33.1	4.9
1993	12	82.6	41.9	13	62.8	25.8	8	41.7	6.8

**Figure 5** Cumulative Acquisitions and Deployments of RDBs



the assimilation gap is the area between the cumulative acquisition and cumulative deployment curves at time *T*, as a proportion of the area under the cumulative acquisition curve at time *T*. Assuming data are captured at discrete intervals, the formula for computing this quantity is:

$$G(T) = \sum_{t=1}^T (A(t) - D(t)) / A(t) \quad (1)$$

where *G(T)* is the estimated assimilation gap at time *T*, *A(t)* is the mean of cumulative acquisitions at times *t*

and *t* - 1, and *D(t)* is the mean of cumulative deployments at times *t* and *t* - 1.

The resulting assimilation gaps for the three technologies are plotted, as a function of time, in Figure 8. To support a more useful comparison across technologies, time is defined as elapsed time since the start year for each, rather than calendar time. The curves are plotted through years 12, 13, and 8, for RDBs, 4GLs, and CASE, respectively; this is the number of years since the start year for each.

The patterns evident in Figure 8 are broadly consistent with the results of the descriptive analysis presented in the previous section. RDBs show the smallest gap, with 54% of the area under the cumulative acquisition curve in year 12 (the last observation year) being accounted for by the area between the acquisition and deployment curves. The estimated RDB assimilation gap is stable over time, ranging only from 0.53 to 0.58 over the last ten observation periods. The gaps for 4GLs and CASE are similar for the first three years after their respective start years, and then they diverge.

### 5.3. Survival Analysis View of the Assimilation Gap

The previous section provides a quantitative measure of the assimilation gap construct according to our primary diffusion modeling concept. In this section we

Figure 6 Cumulative Acquisitions and Deployments of 4GLs

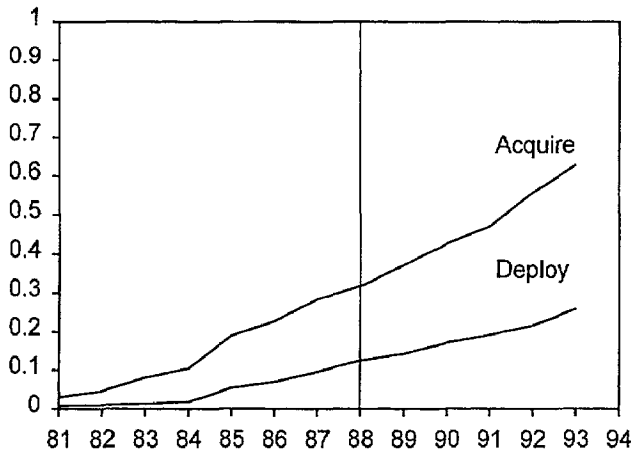
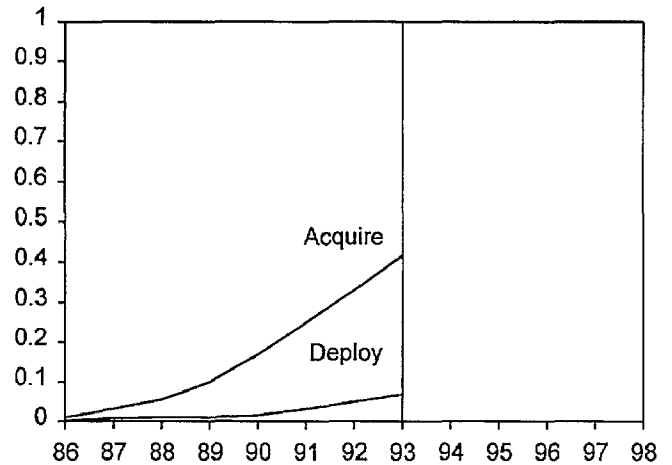


Figure 7 Cumulative Acquisitions and Deployments of CASE

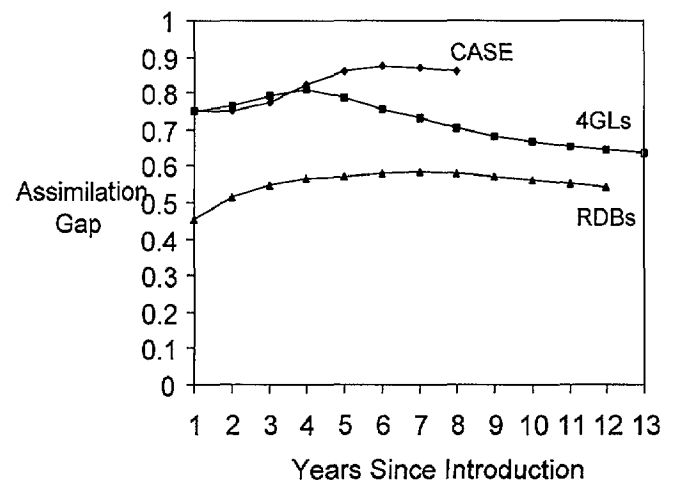


employ survival analysis techniques to further examine the assimilation gaps for these three SPIs. As mentioned earlier, survival analysis techniques not only provide additional insights into the composition of an assimilation gap, but also can be used to support desirable statistical inferences, such as that an assimilation gap is significantly larger (or smaller) than some other gap, or that a particular gap is large enough in absolute terms to be of practical significance taken by itself.

The first step in applying survival analysis is to estimate the population survivor function for each of our three technologies. The population survivor function is defined as the probability that a randomly selected experimental unit from the theoretical population will not have experienced the event of interest by elapsed time  $T$ . Alternatively, the survivor function can be viewed as showing the proportion of the population that will have event time exceeding  $T$ . The sample survivor function is estimated based on the observed event data from a particular sample; if this sample is representative, then the sample survivor function at time  $T$  provides an estimate of the probability that a randomly selected unit from the original population will have a event time greater than  $T$ .

Modeling a survivor function for a particular dataset requires the creation of two variables: an event time variable, and a dummy variable identifying whether a given event time is censored or not. (In this study, a case is censored if the organization has acquired but

Figure 8 Assimilation Gaps



not yet deployed during the observation period.) The time variable contains the reported deployment year minus the reported acquisition year for noncensored cases, and the total observation time for censored cases (calculated as 1993 minus the year of acquisition). With the exception of the first interval, all intervals were defined to be 12 months.<sup>16</sup> The product limit approach, as described by Lawless (1982), was used to estimate

<sup>16</sup>The first interval was defined to be six months because acquisition happens halfway through the reported year of acquisition, on average (assuming a uniform distribution of acquisition times through the year).

the survivor function for each technology.<sup>17</sup> The resulting survivor probabilities at the end of each interval are provided for RDBs, 4GLs, and CASE in Table 2. The survivor functions are plotted in Figure 9.

The survivor functions in Figure 9 provide a view of the assimilation gap that complements the graphically-based measures computed earlier. The more rapidly the survivor function decreases, the smaller the associated assimilation gap will be. While the estimated

<sup>17</sup>This nonparametric approach is preferred here over parametric methods because it requires no assumptions about the functional form for uncensored lifetimes. The Lifestest procedure provided by SAS was used to perform all computations.

**Table 2** Estimated Survival Probabilities for the Acquisition to Deployment Process

Time (months)	RDBs	4GLs	CASE
0	1.0	1.0	1.0
6	0.87	0.87	0.95
18	0.69	0.75	0.88
30	0.53	0.62	0.84
42	0.45	0.60	0.82
54	0.41	0.55	0.76
66	0.37	0.54	
78	0.36	0.50	
90	0.34	0.46	
102	0.27		

survivor functions are based on different lengths of observation windows, the estimation procedure is not sensitive to the length of the data collection period. For example, the survivor estimates produced when only eight time periods were used for RDBs and 4GLs differed by only a few percentage points, on average, compared with those produced when all available data were used (Fichman 1995).

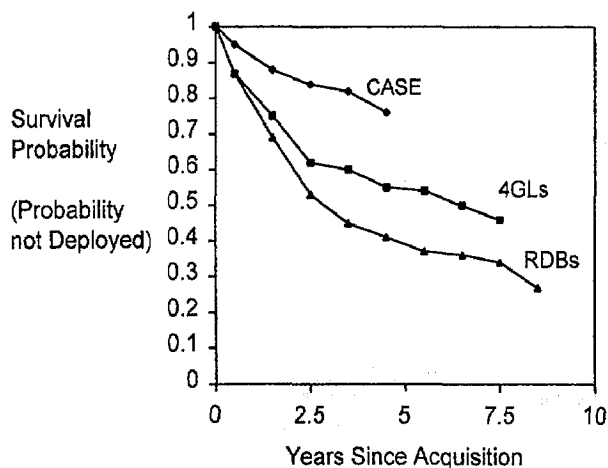
Consistent with the diffusion modeling results, these plots suggest that RDBs have the smallest assimilation gap, with a slightly larger gap for 4GLs, and very pronounced gap for CASE. The survival probabilities (computed using linear interpolation) for RDBs, 4GLs, and CASE at four years are 0.43, 0.57, and 0.79, respectively. The estimated median time to deploy (i.e., the time associated with a survival probability of 0.50) is 3 years for RDBs, 6.5 years for 4GLs, and cannot be determined for CASE, although it appears likely that it will be considerably longer than the value for 4GLs.

The survivor function can also be used to determine a bound for the estimated maximum number of acquirers expected never to deploy. For RDBs, the survival estimate of 0.27 at 8.5 years suggests that a maximum of 27% will never deploy. The survival estimate of 0.46 for 4GLs at 7.5 years likewise suggests that, at most, 46% will never deploy. Since survival estimates could only be calculated through 4.5 years for CASE, the survivor function value at this time (0.76) provides little insight on what the ultimate proportion of CASE nondeployers might be, although a value greater than 50% seems likely. This would mean more than half of CASE tools acquirers will never fully assimilate the tool.

The next step in comparing assimilation gaps is to test whether the observed event times for different technology pairs were likely to have been produced by identical population survivor functions. If not, this suggests that the associated assimilation gaps are significantly different in size. To perform this step, the three individual datasets were concatenated into a single dataset, and a new variable was created to distinguish which cases corresponded to which technology.

In Table 3, the survivor function for each technology is compared with the two other technologies. Two different homogeneity tests were performed for each comparison: the Log Rank test, and the Wilcoxon test

**Figure 9** Survivor Functions for Time to Deployment Since Acquisition



**Table 3** Tests for Homogeneity of Survival Processes

Comparison	Log Rank		Wilcoxon	
	$\chi^2$	<i>p</i> -value	$\chi^2$	<i>p</i> -value
RDB vs. 4GL	7.6	0.006	3.6	0.058
RDB vs. CASE	35.3	0.000	28.3	0.000
4GL vs. CASE	16.4	0.000	15.2	0.000

(see Cox and Oakes 1984).<sup>18</sup> The Log Rank test is more sensitive to differences at larger survival times, while the Wilcoxon is more sensitive to differences at shorter survival times. Five of the six tests reject the hypothesis of homogeneous population survivor functions at  $p \leq 0.05$ , and the sixth test, the Wilcoxon test for homogeneity between RDBs and 4GLs, rejects at  $p \leq 0.1$ .

The results of Table 3 give a high degree of confidence that the assimilation gap for CASE is significantly larger than the ones for both 4GLs and RDBs, and that the one for 4GLs is significantly larger than the one for RDBs. These results are consistent with the conclusions reached earlier, based on visual inspections of cumulative acquisition and deployment curves (Figures 5–7) and based on formalized assimilation gap estimates (Figure 8).

#### 5.4. Comparing an Observed Assimilation Gap to a Baseline Gap

By comparing the survivor functions for two or more technologies, we are able to draw conclusions about the relative sizes of their assimilation gaps. Although this is quite useful on both practical and theoretical grounds, we would also like to be able to compare observed assimilation gaps to some estimated absolute baseline. This can be accomplished by comparing the survivor function for the technology of interest to the survivor function for the baseline. If the two survivor functions are found to be significantly different, and the baseline survivor function lies predominantly below the estimated survivor functions, then this can

<sup>18</sup>A third test, the Log Likelihood Test, is not used here because it assumes an underlying exponential function for uncensored survival times, an assumption not borne out by the data. If an exponential function were appropriate, then the negative log of survival probabilities versus time should produce a linear curve through the origin—but such a pattern does not exist for any of the three SPIs.

serve as a heuristic for concluding that the assimilation gap for the observed process is large enough to be of practical significance. That is, if the observed survivor function can be shown to be significantly “slower” than the baseline function, this would be considered evidence that the assimilation gap is large in an absolute sense.

One way to define the baseline survivor function would be to select some observed technology, i.e., the one with the smallest assimilation gap, and to treat the survivor function for this technology as the baseline. This would suggest using the survivor function for RDBs as the baseline in this study—a conservative approach, since there may be reason to think that the assimilation gap for RDBs is not insignificant. For example, we question whether it seems reasonable that 34% of the acquiring population—the proportion estimated to require more than eight years to deploy RDBs (on just one quarter of new development!)—actually planned to take nearly this long to deploy at the time of acquisition. However, using RDBs as a very conservative baseline would support the conclusion that the assimilation gaps for 4GLs and CASE are large in an absolute sense, owing to the rejection of homogeneity in the statistical tests reported above.

Another approach would be to define the baseline survivor function on normative grounds. For a normative standard for the survivor function, we propose a Weibull<sup>19</sup> distribution in which 75% of acquirers deploy within four years. We believe four years is a generous time for a skilled and appropriate adopter to reach the point of nontrivial deployment for the technologies in this study, and in addition, this allows for 25% of acquirers to be exceptions to this rule. To put this suggested normative baseline in some perspective, Tyre and Orlikowski (1993) report 25 months as the average duration to achieve “full integration” of new

<sup>19</sup>Among the commonly employed distributions for survivor functions, the Weibull is among the most flexible, and is frequently found to be a good fit to observed data (Cox and Oakes 1984). In addition, the log of the negative log of the observed survivor functions were all close to linear, which suggests that the Weibull function is appropriate. The choice of this distribution was further confirmed when the actual fit to the Weibull distribution for all three SPIs was  $R^2 \geq 0.97$ .

capital production equipment among 41 implementation projects. In a study of early adopters of object technology, three of four case study sites achieved a 25% use on new development within four years, despite reporting substantial adoption barriers (Fichman and Kemerer 1997b). It is worth noting that the four year standard was actually achieved by an estimated 57% of the RDBs acquirers in this study (see Table 2). In future research, more well-grounded estimates for the baseline might be obtained by surveying recent acquirers about their deployment plans, by soliciting the opinions of experts (e.g., via Delphi techniques), or by tracking the deployment experiences of a representative set of acquirers.

A five step procedure was used to construct and assess a baseline survivor process conforming to the above definition. Appendix A provides a detailed explanation of this procedure, briefly summarized as follows:

*Step 1:* Specify the parameters (i.e., the "shape" and "rate") of the Weibull survivor function associated with the hypothetical baseline process.

*Step 2:* Use the specified Weibull function to calculate the expected survival probabilities through time  $T$  for the baseline process.

*Step 3:* Use Monte Carlo simulation techniques to randomly generate 30 data sets, each containing survival data conforming to the survival probabilities for the baseline process, as well as the actual survival data for the observed process.

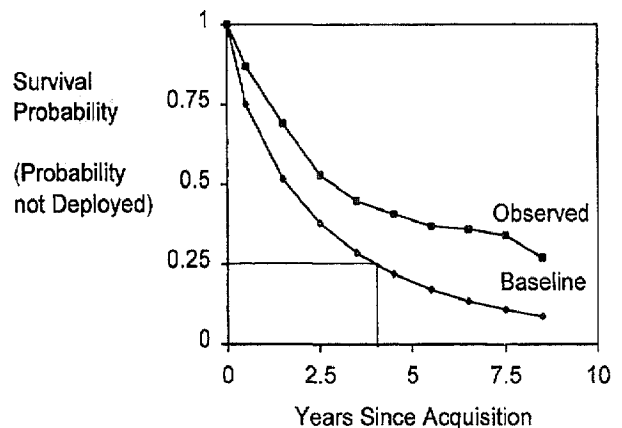
*Step 4:* For each generated data set, perform homogeneity tests to check whether it can be concluded that the baseline series was generated by the same process that generated the observed survivor data.

*Step 5:* Determine whether, in a clear majority of the tests, the conclusion of homogeneity is rejected.

If it turns out that the conclusion of homogeneity is rejected in a clear majority of instances, then this implies that the observed process is significantly different from the hypothetical baseline, and in turn that the assimilation gap for the observed process is large enough to be of practical significance.<sup>20</sup> In Figure 10,

<sup>20</sup>An alternative procedure for comparing an observed survivor function to a baseline is described in Klein and Moeschberger (1997, 6.3). According to this procedure, the analyst computes the differ-

Figure 10 Survivor Functions for RDBs, Actual and Baseline



the survivor function associated with the hypothetical baseline process (i.e., one composed predominantly of ordinary lags) is plotted next to the actual estimated RDB survivor function. In a series of 30 trials, it was concluded in *all* cases that the hypothetical baseline data came from a different process than the observed survival data. This implies that the RDB assimilation gap is significantly larger than the assimilation gap associated with the baseline process. As would be expected, the results were the same for 4GLs and CASE. All 30 trials for each technology supported the conclusion of significant differences from the baseline process.

To determine the sensitivity of these results to the criteria used to define the baseline, a set of three analyses, each containing 30 trials, was performed for each of the three technologies. In the first analysis, the rule for ordinary lags was increased from four to five years. In the second analysis, the rule for the percentage of cases experiencing only ordinary lags was decreased from 75% to 70%. In the third analysis, both of the more

ences between the estimated cumulative hazard functions for the observed technology and for the baseline at time  $T$ , and then compares these estimated differences to the estimated standard errors for those differences. If the 95% confidence interval around these estimate differences does not include zero, then this provides an indirect test of whether the survivor functions are different at time  $T$ . We followed this procedure and found it produced the same inferences as for our more direct simulation approach. These results are available upon request from the authors.

generous values (i.e., 5 years, 70%) were employed simultaneously. Of the 270 total trials, only 5 (all for RDBs using the 5 year, 70% assumption) failed to reject the homogeneity conclusion. These results suggest that while there are significant variations in the size of the gaps among the three technologies, all have gaps that are large enough to be of substantive interest.

## 6. Discussion

In this study, we have developed definitions, measures, and techniques for the study of assimilation gaps, and have demonstrated them in an analysis of three innovations in software process technology. We have shown that substantial assimilation gaps exist, can be sensibly measured, and that their measured size is largely consistent with a priori expectations. All three SPIs exhibited a significant gap, with an especially pronounced gap for CASE. The latter result is consistent with a survey by Howard and Rai that found that among those organizations having acquired any CASE tools, only a negligible number had "substantially replaced conventional systems development techniques with CASE" (Howard and Rai 1993, p. 66).

We have previously argued that the combination of increasing returns and knowledge barriers should especially predispose a technology to exhibit a pronounced assimilation gap (Fichman and Kemerer 1993a). The crux of our argument was that unrealistic expectations—arising from the promise of increasing returns—drive rapid acquisition, but knowledge barriers systematically impede deployment. Therefore, we find it quite interesting that recent reports of CASE tool implementation highlight the role of disappointed expectations and knowledge barriers in discouraging deployment and use. Regarding expectations, Martin notes that "there is a marked gap between CASE promises and actual CASE experiences" (Martin 1995, p. 54). Senn and Wynekoop found that "expectations play a much greater role in the success and failure of CASE implementation than is publicly disclosed. It was repeatedly evident in each study that overly optimistic developer expectations about the tool's relative advantage, even with the top ranked tools used by the organizations studied, increased the likelihood of failure" (Senn and Wynekoop 1995, p. 12). In describing

the implementation of CASE at Amtrak, McComb observed: "Our vendors were quick to tell us how simple it is to learn the new tools. They were quick to tell us we could run the tools on the equipment we had. They promised us they would give us all the support we needed. It wasn't, we couldn't, and they didn't" (McComb 1994, p. 17). Regarding knowledge barriers, Senn and Wynekoop found the complexity of CASE tools to be the number one factor impeding implementation (Senn and Wynekoop 1995). The difficulty of organizational learning surrounding CASE is also emphasized in other recent studies (McComb 1994, Martin 1995, Rai 1995, Guinan et al. 1997).

### 6.1. Immediate Implications

Knowing that a substantial assimilation gap exists—regardless of the precise reason why—can be important information in its own right. When a technology is subject to increasing returns, the rate of arrival of benefits flowing from prior adopters will be an important determinant of whether that technology ever reaches critical mass. Most researchers studying the effects of increasing returns have assumed that adoption is a binary proposition. However, not all kinds of adoption are equal in terms of their contributions to increasing returns. In particular, deployers of an innovation can be expected to contribute to increasing returns in a number of ways that acquirers do not. Deployers buy additional licenses, upgrades, and add-on products and services (thus sustaining economies of scale and learning by doing for producers). They consume complementary products and services that support the core innovation's infrastructure. They often have a useful story to tell other adopters about how to make the innovation work, and contribute to the emerging network of trained employees, user groups, and other network-related resources. Organizations that acquire but do not deploy a technology, by comparison, are likely to make few contributions to increasing returns beyond a one-time addition to producer economies of scale resulting from the initial purchase. Therefore, a pronounced assimilation gap suggests that the benefits of increasing returns—which play a crucial role in sustaining the diffusion of many innovations—may be delayed or even deferred altogether.



As a result, this study suggests that when assimilation gaps are apt to be present, diffusion researchers should use deployment, either instead of, or in addition to, acquisition. In addition, the assimilation gap concept provides a new way to assess the prospects of a technology comparatively early in the overall diffusion cycle. Certain fields, with the information technology field prominently among them, have produced innovations that had extraordinary promise, and initially appeared to be diffusing quite rapidly, but never did approach expected levels of impact and use. This research has shown that it is possible to identify, before the sales peak, that such a pattern may be in evidence. CASE would ordinarily be viewed as quite robust, having been acquired by nearly 50% of the population in only eight years. Yet, as the analysis of assimilation gaps suggests, CASE has barely begun to make its mark in changing the actual practice of software development, and given its slow assimilation, may now be in danger of being superseded by later software process innovations. Technology vendors and mediating institutions can incorporate these insights into their internal studies of the current state and future prospects of emerging technologies. End user organizations can profit from a more sophisticated understanding of what a report of a strongly growing pattern of sales for an innovation says—and does not say—about its future prospects.

## 6.2. Avenues for Theory Testing

The question of why assimilation gaps exist and where they will be most pronounced is itself worthy of further study. Classical notions of diffusion hold that innovation attributes—such as relative advantage, complexity, and compatibility—are relatively fixed, and largely determine the rate and level of diffusion. What, then, is a researcher operating within the confines of classical theory to make of an innovation observed to have diffused rapidly and widely according to one commonly employed definition of the adoption event, but slowly and narrowly according to another common definition? The general explanation articulated here is that the spread of acquisitions and the spread of deployments are, in a sense, different (though related) processes, and are unequally affected by some

important factors. Specifically, we argue that the potential for substantial future benefits created by increasing returns has a stronger positive effect on acquisition than on deployment, while knowledge barriers have a stronger negative effect on deployment than acquisition. When increasing returns and knowledge barriers are both present, as in the case of SPIs, we believe the potential for assimilation gaps becomes very pronounced. These explanations emphasize the dynamic nature of the characteristics of both the innovation and the potential adopter population over time. Although the results of this research are consistent with these explanations, further work would be useful in order to provide systematic confirmation.

A number of avenues are available for such confirmation. One approach would be to capture data on managerial perceptions and intentions at the time of acquisition for a representative sample of firms. Such data could be used, for example, to support or counter the argument that organizations are acquiring an innovation mainly to preserve the option to be able to use it as soon as possible after its benefits have become clear. Alternatively, data could be gathered on the relative prevalence of different adopter level outcomes (excessive delay, implementation failure, stalling, discontinuance, etc.) and managerial perceptions of the factors leading to those outcomes. These kinds of data could be used to confirm or deny arguments that attribute the gap to deployment difficulties or disappointed expectations arising from knowledge barriers and/or immaturity of the technological network. A third approach would be to gather data on adopter characteristics, such as those related to innovative capabilities or the competitiveness of the external business environment. These data could then be used to stratify the sample and statistically test whether some adopter groups have larger assimilation gaps than others. Larger assimilation gaps among those with lesser innovative capabilities would support theories based on knowledge barriers. Larger gaps for adopters residing in firms with stronger competition would point towards options-based arguments. A fourth approach would be to gather, for several innovations, data on perceived innovation characteristics (such as implementation complexity, technical potential, extent of signaling) at the point when the technologies are first

introduced, and to assess the extent to which these variables correlate with measured assimilation gaps for those innovations.

### 6.3. Other Explanations for the Assimilation Gap

While we have focused on the role of increasing returns and knowledge barriers in producing an assimilation gap, two other classes of explanations are worth mentioning, both predicated on the notion that for some technologies adoption involves two separate decision processes, one to acquire the innovation and one to deploy it (Leonard-Barton 1988).

The first argument is that the existence of multiple decisions opens the possibility of different decision makers, with correspondingly different motivations and decision criteria. Under this scenario, more senior decision makers might systematically favor acquisition of a technology based on its expected benefits at the level of the business unit, or because of the symbolic benefits of being viewed as an innovator (Feldman and March 1981), or because of institutional pressures (Rowan 1977, Abrahamson 1996)—only to subsequently encounter pervasive resistance among those that must actually deploy or use innovation. This resistance might arise because the innovation is viewed by most secondary adopters as too risky (Cooper and Zmud 1990), deskilling (Attewell and Rule 1984), competence destroying (Tushman and Anderson 1986), incompatible (Ramiller 1994), or in some other way undesirable. Under this argument, then, the assimilation gap arises from a discrepancy between the knowledge and motivations of those responsible for acquisition versus those responsible for deployment and use.

The second argument is that even if the same decision makers are involved in both the acquisition and deployment decisions, different considerations might be salient for each decision. Abrahamson (1996) has argued, for example, that many innovative management ideas are subject to the same forces that drive fad and fashion, and that managers may be compelled to adopt simply owing to popularity of the ideas among members of their institutional network. While Abrahamson does not make a distinction between acquisition and deployment, it seems possible that fad and fashion might compel the outward appearance of adoption created by acquisition, but not similarly compel actual use.

### 6.4. Potential Limitations

Like most studies, this one is subject to some limitations. In terms of data collection, managers were asked to provide retrospective reports on the timing of events that, in some cases, occurred many years in the past. This can be expected to introduce noise into the measurement process. However, we are unaware of any reason to think it introduces systematic bias. Second, the definitions for nontrivial deployment (25% use on new applications), the duration for ordinary lags (4 years), and the baseline assimilation gap (75% experiencing only ordinary lags), while designed to be generous standards for the technologies included in this research, are somewhat arbitrary and might be too strict (or not strict enough) for other technologies. However, a sensitivity analysis using different values for the latter two rules had no effect on the results. Finally, for CASE, the portion of the diffusion process encompassed by the observation period was less than might ideally be desired, but unavoidable given the relative newness of the technology. For all three innovations, only the portion of the innovation process for which both acquisitions and deployments were still experiencing steady growth was observed, thus accounting for the relative stability of measured gaps over time. Nevertheless, these potential limitations are not believed sufficient to bring into serious question the broad empirical conclusions of the study, namely that assimilation gaps exist, are measurable, and are large enough, especially for CASE, to merit practical attention.

## 7. Conclusions

This research makes several contributions to the study of innovation diffusion. It introduces the assimilation gap concept, and develops a general operational measure derived from the difference between the patterns of cumulative acquisitions and deployments. It presents several theoretical explanations for why some technologies may be especially predisposed to a pronounced gap. It develops novel techniques for measuring assimilation gaps, for establishing whether two gaps are significantly different from each other, and for establishing whether a particular gap is absolutely large enough to be of substantive interest. Finally, it

demonstrates these techniques in an analysis of adoption data for three prominent innovations in software process technology—RDBs, 4GLs, and CASE. The analysis confirmed that assimilation gaps can be sensibly measured, and that their measured size is largely consistent with a priori expectations. As expected, all three SPIs exhibited significant assimilation gaps. The gap for CASE was especially pronounced, a result consistent with the growing chorus of studies noting marked CASE implementation challenges.

Assimilation gaps imply a distinctive pattern of diffusion, one that has important implications for managers and researchers. An innovation that enjoys robust sales—yet is only sparsely deployed—is not genuinely diffusing in the sense of having a significant impact on the operational processes of acquiring firms. This suggests that based on impressive sales data alone, observers should be guarded about concluding that an innovation is necessarily destined to become widely used. For diffusion researchers, this study has the immediate implication that cumulative deployments should be modeled instead of, or in addition to, cumulative acquisitions when assimilation gaps are likely to be present. In addition, this work lays a foundation of concepts and techniques researchers can use in future theory-driven investigations of why assimilation gaps exist, and what can be done to predict and/or counter them.

The authors gratefully acknowledge the financial support provided for this research by the MIT Center for Information Systems Research (CISR). We also thank Sue Conger, Randy Cooper, Kathy Eisenhardt, three anonymous reviewers, and the associate editor for their helpful comments.

### Appendix A: Using the Weibull Distribution to Construct and Analyze Data for a Hypothetical Baseline Survivor Process

*Step 1.* Determine the “shape” and “rate” parameters for the baseline survivor function.

The Weibull survivor distribution (see Equation (2) below), has two parameters,  $\kappa$  and  $\rho$ . The  $\kappa$  parameter has the greatest influence on the shape of the curve, while  $\rho$  has the most influence on the rate at which it declines

$$S(t) = e^{-(\rho t)^\kappa} \quad (2)$$

To define the baseline survivor process, both the shape and the rate parameter for the associated Weibull distribution must be determined. This was done by assigning the shape parameter based

on the shape of the *observed* survivor process to which the baseline will be compared, and then assigning the rate parameter to ensure that the baseline process behaves as intended, i.e., it passes through the point of 75% deployment at 48 months.

To obtain a value for  $\kappa$  for each baseline process, least squares regression (after appropriate transformations) was used to fit a Weibull distribution to the actual survivor function estimates for the associated technology.  $\kappa$  for the baseline ( $\kappa'$ ) was then assigned to be the same as the estimated  $\kappa$  parameter for the associated observed technology.  $\rho$  for the baseline ( $\rho'$ ) was determined by simply constraining the baseline curve to pass through 0.25 at 48 months. That is, the equation  $S(t) = 0.25 = e^{-(\rho' \cdot 48)^{\kappa'}}$  was solved for  $\rho$ .

*Step 2.* Use the specified Weibull function to calculate the expected survival probabilities through time  $T$  for the baseline process.

This was accomplished by simply solving the survivor function equation for each technology repeatedly for  $T = 6, 18, 30$  and so on until the maximum possible observed duration was reached.

*Step 3.* Use Monte Carlo simulation techniques to randomly generate multiple data sets, each containing survival data conforming to the survival probabilities for the baseline process and the observed process.

Once survival probabilities have been calculated, a random number generator can be used to generate survival times consistent with these assumed probabilities. (The probability of a survival time falling within any particular interval is equal to survival probability at the end of the interval less the survival probability at the beginning of the interval.) The specific procedure employed was to randomly assign, for every case where the acquisition of a particular technology had been reported, a random time until deployment for that case. For every deployment time that resulted in a deployment year later than 1993, the case was designated as being censored.

*Step 4.* For each generated dataset, perform tests to check whether the hypothesis that the series was generated by the same process that generated the observed survivor data can be rejected.

Homogeneity tests were performed comparing the actual process to the associated baseline just as they were to compare two actual survival processes.

*Step 5.* Determine whether, in the clear majority of the trials, the conclusion of homogeneity is rejected.

This was accomplished by running 30 simulations for each technology and recording the results of the homogeneity tests for each.

### References

- Abrahamson, E. 1996. Management fashion. *Acad. Management Review* 21(1)254–285.
- Arrow, K. 1962. The economic implications of learning by doing. *Review Econom. Studies* 29:166–170.
- Arthur, W. B. 1988. Competing technologies: an overview. G. Dosi, C. Freeman, R. Nelson, G. Silverberg, L. Soete, eds. *Technical Change and Economic Theory*. Pinter Publishers, London, 590–607.
- . 1996. Increasing returns and the new world of business. *Harvard Bus. Rev.* 74 (July–August) 101–109.
- Attewell, P. 1992. Technology diffusion and organizational learning: the case of business computing. *Organ. Sci.* 3(1)1–19.

- . J. Rule. 1984. Computing and organizations: what we know and what we don't know. *Comm. ACM* 27 (December) 1184–1192.
- Bass, F. M. 1969. A new product growth model for consumer durables. *Management Sci.* 15(1)215–227.
- Brancheau, J. C., J. C. Wetherbe. 1990. The adoption of spreadsheet software: testing innovation diffusion theory in the context of end-user computing. *Inform. Systems Res.* 1(2)115–143.
- Brynjolfsson, E., C. F. Kemerer. 1997. Network externalities in microcomputer software: an econometric analysis of the spreadsheet market. *Management Sci.* 42(12)1627–1647.
- Cohen, W. M., D. A. Levinthal. 1990. Absorptive capacity: a new perspective on learning and innovation. *Admin. Sci. Quart.* 35(March)128–152.
- Cooper, R. B., R. W. Zmud. 1990. Information technology implementation research: a technological diffusion approach. *Management Sci.* 36(2)123–139.
- Cox, D. R., D. Oakes. 1984. *Analysis of Survival Data*. Chapman and Hall, London, UK.
- Dillman, D. 1983. *Mail and Telephone Surveys*. John Wiley & Sons, New York.
- Eveland, J. D., L. G. Tornatzky. 1990. The deployment of technology. L. G. Tornatzky, M. Fleischer, eds. *The Processes of Technological Innovation*. Lexington Books, Lexington, MA, 117–148.
- Feldman, M. S., J. G. March. 1981. Information in organizations as signal and symbol. *Admin. Sci. Quart.* 26171–186.
- Fichman, R. G. 1992. Information Technology Diffusion: A Review of Empirical Research. *Proc. Thirteenth International Conf. on Information Systems*, Association for Computing Machinery Dallas, TX.
- . 1995. The assimilation and diffusion of software process innovations. Unpublished PhD thesis, MIT Sloan School of Management, Cambridge, MA.
- , C. F. Kemerer. 1993a. Toward a theory of the adoption and diffusion of software process innovations. L. Levine, ed. *Diffusion, Transfer and Implementation of Information Technology*, IFIP TC8 Working Group Conference Proceedings, Champion, PA. Elsevier Science Publications, New York, 23–31.
- , —. 1993b. Adoption of software engineering process innovations: the case of object orientation. *Sloan Management Rev.* 34(2)7–22.
- , —. 1997a. The assimilation of software process innovations: an organizational learning perspective. *Management Sci.* 43(10)1345–1363.
- , —. 1997b. Object technology and reuse: lessons from early adopters. *IEEE Computer* 30(10)47–59.
- Guinan, P. J., J. G. Coopridge, S. Sawyer. 1997. The effective use of automated application development tools. *IBM Systems J.* 36(1)124–139.
- Gurbaxani, V. 1990. Diffusion in computing networks: the case of BITNET. *Comm. ACM* 33(12)65–75.
- , H. Mendelson. 1990. An integrative model of information systems spending growth. *Inform. Systems Res.* 1(1)23–47.
- Howard, G. S., A. Rai. 1993. Promise and problems: CASE usage in the US. *J. Inform. Technology* 8(2) 65–73.
- Katz, M. L., C. Shapiro. 1986. Technology adoption in the presence of network externalities. *J. Political Econom.* 94(4)822–841.
- Kemerer, C. F. 1992. How the learning curve affects CASE tool adoption. *IEEE Software* 9(3) 23–28.
- Klein, J. P., M. L. Moeschberger. 1997. *Survival Analysis: Techniques for Censored and Truncated Data*. Springer, New York.
- Kogut, B., U. Zander. 1992. Knowledge of the firm, combinative capabilities, and the replication of technology. *Org. Sci.* 3(3)383–397.
- Kwon, T. H., R. W. Zmud. 1987. Unifying the Fragmented Models of Information Systems Implementation. J. R. Boland, R. Hirschheim, eds. *Critical Issues in Information Systems Research*. John Wiley, New York, 227–251.
- Lawless, J. F. 1982. *Statistical Models for Lifetime Data*. John Wiley & Sons, New York.
- Leonard-Barton, D. 1988. Implementation characteristics of organizational innovations. *Comm. Res.* 15(5)603–631.
- . 1989. Implementing new production technologies: exercises in corporate learning. M. A. Von Glinow, S. Mohrman, eds. *Managing Complexity in High Technology Industries: Systems and People*. Oxford University Press, New York.
- Liker, J. K., M. Fleischer, D. Arnsdorf. 1992. Fulfilling the promises of CAD. *Sloan Management Rev.* 33(3) 74–86.
- Mahajan, V., E. Muller, F. M. Bass. 1990. New product diffusion models in marketing: a review and directions for research. *J. Marketing* 54 (January)1–26.
- , R. Peterson. 1985. *Models for Innovation Diffusion*. Sage Publications, Beverly Hills, CA.
- Mansfield, E. D. 1993. The diffusion of flexible manufacturing systems in Japan, Europe and the United States. *Management Sci.* 39(2)149–159.
- Markus, M. L. 1987. Toward a 'critical mass' theory of interactive media: universal access, interdependence and diffusion. *Comm. Res.* 14(5)491–511.
- Martin, M. P. 1995. The case against CASE. *J. Systems Management* 46(1)54–57.
- McComb, M. E. 1994. CASE tools implementation at Amtrak—Lessons almost learned. *J. Systems Management* 45(3)16–20.
- McGrath, R. G. 1997. A real options logic for initiating technology positioning investments. *Acad. Management Rev.* 22(4)974–996.
- Meyer, A. D., J. B. Goes. 1988. Organizational assimilation of innovations: a multilevel contextual analysis. *Acad. Management J.* 31(4)897–923.
- Orlikowski, W. J. 1993. CASE tools as organizational change: investigating incremental and radical changes in systems development. *MIS Quar.* 17(3)309–340.
- Prescott, M. B. 1995. Diffusion of innovation theory: borrowings, extensions, and modifications from IT researchers. *Database* 26(2,3)16–19.
- , S. A. Conger. 1995. Information technology innovations: a classification by IT locus of impact and research approach. *Database* 26(2,3)20–41.
- Rai, A. 1995. External information source and channel effectiveness and the diffusion of CASE innovations: an empirical study. *European J. Inform. Systems* 4 93–102.

- Ramiller, N. C. 1994. Perceived compatibility of information technology innovations among secondary adopters: toward a reassessment. *J. Engrg. Tech. Management* 11 1-23.
- Rogers, E. M. 1995. *Diffusion of Innovations*. The Free Press, New York.
- Rosenberg, N. 1976. On technological expectations. *Econom. J.* 86 (September) 523-535.
- . 1982. *Inside the Black Box: Technology and Economics*. Cambridge University Press, Cambridge, U.K.
- Rowan, B. 1977. Organizational structure and the institutional environment. *Admin. Sci. Quart.* 27:259-279.
- Saltzman, A. 1993. Improving response rates in disk-by-mail surveys. *Marketing Res.* 5 (3)32-39.
- Schilling, M. A. 1998. Technological lockout: an integrative model of the economic and strategic factors driving technology success and failure. *Acad. Management Rev.* 23 (2)267-284.
- Senn, J. A., J. L. Wynekoop. 1995. The other side of CASE implementation: best practices for success. *Inform. Systems Management* 12 (4)7-14.
- Singer, J. D., J. B. Willett. 1991. Modeling the days of our lives: using survival analysis when designing and analyzing longitudinal studies of duration and timing of events. *Psych. Bull.* 110 (2)268-290.
- Stone, J. A. 1993. *Inside ADW and IEF: The Promise and Reality of CASE*. McGraw-Hill, New York.
- Swanson, E. B. 1994. Information systems innovations among organizations. *Management Sci.* 40(9) 1069-92.
- Tornatzky, L. G., M. Fleischer eds. 1990. *The Processes of Technological Innovation*. Lexington Books, Lexington, MA.
- Tushman, M. L., P. Anderson. 1986. Technological discontinuities and organizational environments. *Admin. Sci. Quart.* 31 439-465.
- Tyre, M. J., W. J. Orlikowski. 1993. Exploiting opportunities for technological improvement. *Sloan Management Rev.* 35 (1)13-26.
- Van de Ven, A. H. 1993. A community perspective on the emergence of innovations. *J. Engrg. and Tech. Management* 10:23-51.
- von Hippel, E. 1994. Sticky Information' and the locus of problem solving: implications for innovation. *Management Sci.* 40 (4)429-439.

Michael Vitale, Associate Editor. This paper was received on October 24, 1997 and has been with the authors 1 year for 1 revision.