12-13-2015

Demystifying Big Data Adoption: Beyond IT Fashion and Relative Advantage

Hong-Mei Chen
University of Hawaii at Manoa, hmchen@hawaii.edu

Rick Kazman
University of Hawaii at Manoa, kazman@hawaii.edu

Florian Matthes
Technical University of Munich, matthes@tum.de

Demystifying Big Data Adoption: Beyond IT Fashion and Relative Advantage

Completed Research Paper

Hong-Mei Chen
University of Hawaii at Manoa
hmchen@hawaii.edu

Rick Kazman
University of Hawaii at
Manoa Kazman@hawaii.edu

Florian Matthes
Technical University of Munich
matthes@tum.de

Abstract

There is a paradox in big data adoption: a peak of hype and simultaneously an unexpectedly low deployment rate. The present multiple case study research develops a Big Data Adoption (Big²) model that helps to explain this paradox and sheds light on the "whether", "why", and "how" questions regarding big data adoption. The Big² model extends beyond the existing Relative Advantage and IT Fashion theories to include organizational, environmental, social variables as well as new psychological factors that are unique to big data adoption. Our analysis reveals that the outcome of big data adoption is indeterministic, which defies the implicit assumption of most simplistic "rational-calculus" models of innovation adoption: Relative Advantage is a necessary but not sufficient condition for big data adoption. Most importantly, our study uncovered a "Deployment Gap" and a "Limbo Stage" where companies continuously experiment for a long time and do not proceed to deployment despite the intent to adopt big data. As a result there are four big data adoption categories: Not adopting, Experimented but Not Adopting, Not Yet Deployed, Deployed. Our Big² model contributes to provide a Paradigm Shift and Complexity Tolerance perspective to understand the "why" in each of the 4 adoption categories. This study further identifies 9 complexity tolerance strategies to help narrow the Deployment Gap but also shows that big data is not for everyone.

Keywords: Big Data Adoption model, Enterprise Emerging IT Innovation Adoption, Firm-level adoption factors, Deployment Gap, Complexity Tolerance

Demystifying Big Data Adoption: Beyond IT Fashion and Relative Advantage

Introduction

Big data is possibly the most significant technology disruption since the meteoric rise of the Internet and the digital economy (Agarwal and Dhar 2014). MGI predicts that big data will become a key basis for competition, productivity growth, and innovation (MGI, 2011). However, despite being at the peak of the 2014 Hype Cycle for Emerging Technologies (Gartner 2014a), Gartner (2014b) reported "2013 is the year of big data experimentation, 2014 too." At the end of 2014, big data deployments are scarce (Gartner 2014b): only 13% of respondents said they put big data projects into production in 2014, and failures abound. In 2013, 55% of big data projects were not completed (Infochimps 2013).

"Why is there such high hype but low adoption rates and few deployments?" we wondered. This is a paradox, different from previous enterprise IT adoption patterns, such as CRM (Customer Relationship Management), ERP (Enterprise Resource Planning), SOA (Service Oriented Architecture), etc. Enterprise Adoption of innovations has been a topic of significant interest for decades as enterprises increasingly depend on IT innovation to grow, compete, or survive. To help practitioners shape effective business strategies and ensure competitive advantage, it is essential to understand why, how, and under what conditions enterprises have succeeded or failed in adopting IT innovations. However, despite the large number of innovation adoption studies, the majority of research has focused on well-understood IT innovations. Little work has examined emerging IT innovations such as big data, and scant empirical evidence has been provided (Oliveira and Martins, 2011; Basole et al. 2013; Orlando et al. 2013; Baskerville et al. 2014).

The present empirical study seeks to understand the enterprise adoption of an emerging enterprise IT innovation: big data, which became a mega-trend only recently, since 2011, when IBM created the #bigdata tag (Winshuttle.com 2014). Our research questions center on exploring the "why" of the paradoxical big data adoption phenomena: Why do enterprises (intend to) adopt or not (intend to) adopt big data? What factors drive their decisions? Our subsequent research questions are on assisting enterprises on the "whether" and "how" to adopt big data.

On "why" organizations adopt IT innovation, existing firm-level adoption theories include Diffusion of Innovation (DOI) (Rogers, 2003) Technology-Organizational-Environment (TOE) (Tornatzky & Fleischer, 1990) and more recently IT Fashion theory (Wang, 2010). The "Relative Advantage" variable of DOI has been extensively used to explain adoption: decision markers rationally compare the advantages of adopting new technology relative to their existing systems, an economic–rationalistic perspective. The TOE framework extends DOI theory by including organizational and environmental contexts of the innovation adoption. IT Fashion theory, rooted in institutional theory, extends DOI theory by considering the social setting of emerging IT innovation. Abahamson (1996) defines a social group—fashion setters such as consultants, journalists and academics—who promote innovations as 'must-do' or 'best-practices,' i.e., the hype. Hype attracts fierce enthusiasm and attention and triggers high expectations of benefits. Previous widely adopted IT innovations such as data warehouse, ERP, and CRM frequently go through such fashion periods (Wang, 2010). They underwent wide swings in popularity, progressed through hype cycles filled with inflated expectations, and spread across organizations via passionate bandwagons of adoption (Wang, 2010).

However, big data adoption does not follow the same path. Why? What factors play a role in the paradoxical phenomena of big data adoption? In fact, the findings from 11 year (2000 to 2011) (Orlando et al, 2013) and 31 year (1977 to 2008) (Basole et al. 2013) literature surveys indicate that there is an extremely varied approach to the recognition of organizational level adoption determinants and there appears to be no consensus. As existing theories are not able to explain the paradoxical phenomena of big data adoption, we conducted an empirical multiple case study of 25 European enterprises, utilizing a grounded theory method (Glaser and Strauss 1967; Glaser 1992). Our study identified 9 new big data adoption factors that go beyond

Relative Advantage and IT fashion, forming the Big Data Adoption ($Big^{2)}$ Model. In explaining the paradoxes of big data adoption, we uncovered a "Deployment Gap" and a "Limbo Stage" where companies continuously experiment for a long time and do not proceed to deployment despite the intent to adopt big data. Based on these observations, we have defined four big data adoption categories. Our results reveal IT fashion has different effects on each of the 4 categories. Most importantly, our results show that the characteristics and paradigm shifts associated with big data increase the complexity needed to be "tolerated" by enterprises for moving to deployment. The Complexity Tolerance variable is a result of the interplay between the rational and non-rational (psychological) factors. In what follows, we will first describe the characteristics of big data and paradigm shifts to understand the layered complexity and high level of uncertainly and risk involved with big data adoption and how it is different from previous fashioned IT.

Big Data 5V Characteristics and Adoption Challenges

What is big data? Many definitions exist. The 1st V of big data is **Volume** (scale of data): "data whose size forces us to look beyond the tried-and true methods that are prevalent at that time." (Jacobs 2009). The 2nd V is **Variety** (different forms of data). "Big data is a buzzword, or catch-phrase, used to describe a massive volume of both structured and unstructured data that is so large that it's difficult to process using traditional database and software techniques." (Beal 2014). The 3rd V is **Velocity** (analysis of streaming data): "Big Data consists of extensive datasets, primarily in the characteristics of volume, velocity, and/or variety that require a scalable architecture for efficient storage, manipulation, and analysis" (NIST 2014). IBM (2012) introduced the 4th V, **Veracity** (the uncertainty of data or uncertain quality of data), as data are collected from everywhere including web logs, IoT (Internet of Things), social media, mobile data, etc. and it may include so called "dirty data." And finally the 5th V of big data is **Value**. Much value is expected to be derived: big data has the potential to fundamentally transform organizational processes, business models and strategies, and even entire industries and markets. Managing, analyzing, visualizing, and extracting useful information from large data sets will help organizations increase operational efficiency, inform strategic direction, develop new products and services, identify new customers and markets, make better decisions, and become more innovative. (Davenport, Barth, & Bean 2012).

However, the *value* of big data is subject to debate. Enthusiasts claim that the data deluge makes the scientific method obsolete. "With enough data, the numbers speak for themselves." (Anderson 2008). The success of Google's behavioral targeting strategies enabled by big data supports this view. However, critics caution hidden biases of big data (Crawford 2013) and describe it as a cultural, technological, and scholarly phenomenon that rests on the interplay of technology, analysis and *mythology*—the widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with an aura of truth, objectivity, and accuracy (Boyd and Crawford 2012). Big Data and whole data are also not the same. Bigger data is not always better data. Without taking into account the *sample*, the size of the data set is meaningless. (Boyd and Crawford 2011). Big Data is a powerful tool for inferring correlations, not a magic wand for inferring causality. Google Flu Trends' failure to predict peak flu levels highlights the debates about big data value (Dodson, 2014).

While the debate of big data value is ongoing, an array of technology has proliferated and evolved rapidly to address the technical challenges stemming from the 5Vs (Chen et al. 2015). This adds a new dimension of complexity to system development and thus complicates enterprise adoption decisions. The adoption of big data involves selection of many technology components to **orchestrate** a big data system. Currently, big data technologies include distributed database processing, MPP (Massively Parallel Processing) databases, NoSQL databases, New SQL databases, in-memory databases, real-time stream processing, advanced analytics, big data analytic clouds, and big data appliances, just to name a few. Each technology type has many vendors and products. This large space of technology choices is problematic (Chen et al. 2015). The challenges in technology selection are non-trivial and affect the choices of architecture patterns, data models, programming languages, query languages, and access methods. These choices, in turn, affect performance, latency, scalability, availability, consistency, modifiability, etc. Furthermore, many new technologies such as data lake, enterprise data hub, data refinery, Lambda architecture (Marz and Warren 2013) and Polyglot persistence (Sadalage and Fowler 2013) have recently appeared. And the integration of these new systems with existing ones is a challenge for companies who cannot afford to start anew (Chen et al. 2015).

The technology selection, orchestration and integration challenges are *only* one of the many system development paradigm shifts induced by the 5Vs of big data that the enterprises will have to grapple to have a successful adoption. Big data adoption decisions involve layered complexity and high levels of uncertainly and risk. These potentially involve drastic structural and process changes in the organization, new requirements for system development, and major resource allocation decisions. And as the costs and benefits of big data adoption cannot be known *a priori* due to the scope and scale of big data and the speed of technology changes. The outcome of adoption is *indeterministic* as the value of big data depends on how it is implemented and the interplay of many technical, organizational, political and environmental issues (Chen et al. 2010). All these defy the implicit assumption of most simplistic "rational-calculus" models of innovation adoption. If the consequences of the innovation can be clearly understood, then the decision-making process can be a purely rational one, derived solely on the efficiency drivers that are associated with the innovation (Rogers 2003). This is not in the case for big data.

Research Method

We conducted a multiple case study of 25 European enterprises to help shed light on the paradoxes underlying big data adoption. Case study research is deemed suitable when the proposed research addresses a contemporary phenomenon which the researcher has no control over, the research is largely exploratory and addresses the "how" and "why" questions (Yin 2014). Multiple cases are suggested to increase the methodological rigor of the study, particularly, because "evidence from multiple cases is often considered more compelling" (Yin 2014).

In addition, we utilized a grounded theory method for our exploratory study. We followed the steps of grounded theory outlined in (Sbaraini et al. 2011). Our research process began with broad research questions: we did not have *a priori* identification of variables or preconceived hypotheses, but we did have a broad understanding of the literature and theories in many IS subfields. The grounded theory method is particularly appropriate for studying emerging phenomena that go beyond the context of existing theories.

Case Selection

Our case selection was partly opportunistic, based on accessibility. We focused on large enterprises (> 5,000 employees) that, we believed, had the resources to bear the high upfront costs of big data investment. We also contacted outsourcing companies because the professional services category is the largest segment of big data market (Kelly 2014). We had two sources of contacts:

- Source 1: the TUM Sebis (Software Engineering for Business Information Systems) database containing more than 700 contacts in industry. We selected 55 individuals in 33 enterprises to contact. 35 individuals did not response; 18 individuals from 18 organizations agreed to participate in the research; 2 responded as unable to participate, citing time constraints.
- Source 2: 7 organizations with which the researchers had prior collaborations.

From these 2 sources, we obtained a total of 25 cases: 23 large enterprises, 1 medium, 1 small. The average employee size is > 150,000. Several enterprises belong to 1 conglomerate. There are 5 outsourcers and 20 non-outsourcer enterprises. 21 were German-based, and 4 were in Europe but outside Germany. 1 of the 4 non-German companies was headquartered in the U.S. The industries include: telecommunication services; manufacturer: automation & power; manufacturer: airplanes; global financial services; logistics; airline; reinsurance and financial services; manufacturer: smart plants; conglomerate; financial services; insurance and reinsurance; manufacturer: automotive components; manufacturer: automotive; investment banking; utility: energy, telecom, and IT; insurance software provider; tax and legal software provider; general outsourcing; outsourcer: energy technology provider; outsourcer: smart city and energy; outsourcer: telecom IT services. With 25 organizations, we reached a *saturation point* where we no longer heard anything new and believed that more companies and data would not change the emerging theory. 25 cases is considered a large sample size in grounded theory.

Data Collection

We collected data from multiple sources including public corporate information, management consultant reports, magazine and newspaper articles, informal exchanges with colleagues, formal interviews, site visits, documents (slides shows, internal technical reports, use cases for big data, etc.) given by companies, and collaborative practice research (CPR) (Mathiasen 2002) with 2 outsourcers. One author worked for one conglomerate. We conducted 28 formal semi-structured interviews with 40 individuals, typically lasting 1 to 3 hours, not including pre-interview or follow-up emails and phone conversations. Our CPR work (Chen et al. 2015) sessions lasted many hours during collaboration periods. We also had additional informal conversations where questions were discussed. Interviewees in large enterprises included C-level executives, VPs, Division Heads, Chief Architects, and Enterprise Architects. 84% of interviewees in this category had a Ph.D. and 3 interviewees won a CIO award in Germany. Interviewees in outsourcing companies included CEO, VP, R&D Division head, chief architect, and lead programmers. Only 3 out of 40 interviewees were female.

Data Coding and Analysis

We designed an Interview Question Catalog (IQC) containing potential questions for use in the formal semi-structured interviews. We studied the companies and the interviewees' publicly available information to maximize the efficient use of interview time. However, we kept our questions open for interviewees to freely speak about what they deemed important. The interview questions and process were dynamically adjusted to fit the interviewee's interaction styles. We took extensive notes but did not otherwise record the conversations to minimize disruptions. Immediately after each interview we compared the notes from both interviewers for accuracy and completeness. We wrote "memo" and draft model-like graphs for conceptualizing what we discovered. In these memos, we made comparisons between data, cases and codes to find similarities and differences, and raised questions to be answered in continuing interviews. We coded and analyzed field memos, interview notes (and messages exchanged at other times, if applicable) to derive concepts. We categorized and organized the concepts that emerged in interviews and other interactions. We linked (when appropriate) the concepts that emerged in one interview to concepts from prior interactions. We prepared new questions generated from the concepts that emerged for the next round of interactions. The IQC was constantly updated and we constructed a Project Database for storing and managing all research records including correspondence, phone conversations, company profiles, financial information, industry studies, reports from various sources, memos, interviews, observations, site visits, documents given by the companies, etc.

Results

Before we present the resulting model, we want to clarify that the focus of our study is on adoption, not diffusion. Adoption is the introduction of something new to an organization while diffusion refers to the process by which an innovation spreads among organizations (Frambach and Schillewaert 2002; Rogers 2003; Zmud 1982). The diffusion process begins with introduction of the innovation to the population, and ends when the population is saturated, i.e., when all those who will ever adopt have, in fact, adopted (Fichman, 2004a). The diffusion aspect has been excluded from consideration in our study as it refers to mechanisms and events that take place *after* innovation adoption (Orlando 2013).

The adoption process has three major phases: initiation, adoption (decision) and implementation (Rogers 2003). Initiation refers to the organization's initial awareness and consideration of the innovation. Later, there is the adoption decision, which involves evaluation of the innovation from technical, financial and strategic perspectives (Damanpour and Schneider 2006). Finally, the implementation stage occurs when the organization purchases the innovation and prepares to use it, until it becomes a routine feature (Damanpour and Schneider 2006; Rogers 2003). Our study focus on the initiation and adoption (decision) stages leading to deployment, e.g., putting the big data project into production.

Big Data Adoption Process

As shown in Figure 1, our study found a distinct process in big data adoption. Our case enterprises typically initiate the consideration of big data adoption through an organizational innovation process. The innovation process could be top down, bottom up, or a mix of both. In our study, due to the size of our enterprises, there typically exists a formal corporate process for introducing the consideration of emerging technologies. A typical top-down innovation process involves upper executive support, setting up a steering committee, scheduled meetings, communications with departments to request inputs, etc. A bottom-up innovation process is initiated by a line of business (LOB) desiring new capabilities from big data. In most cases, if the attitude to big data is favorable, the intention to adopt may be revealed in the Innovation Process stage. An intensive Value Exploration stage contains a series of assessments and high level demonstrations of the value of big data. In the Use Case Development stage, the possible "use cases" are defined to determine how the technology can provide value. In this stage, various LOB will be asked to brainstorm big data scenarios and submit these to the steering committee. From the use cases collected (typically hundreds), a few will be selected to be prototyped in the Experiment stage. After the Experiment stage, a decision to adopt will then lead to Deployment Planning, deciding how the system will be implemented. Outsourcing agreements and new hires usually occur in this stage. In the Deployment stage, the deployment plan is put into action. (Note that deployment here is not the same as deployment in the software development process where it refers a software release.)

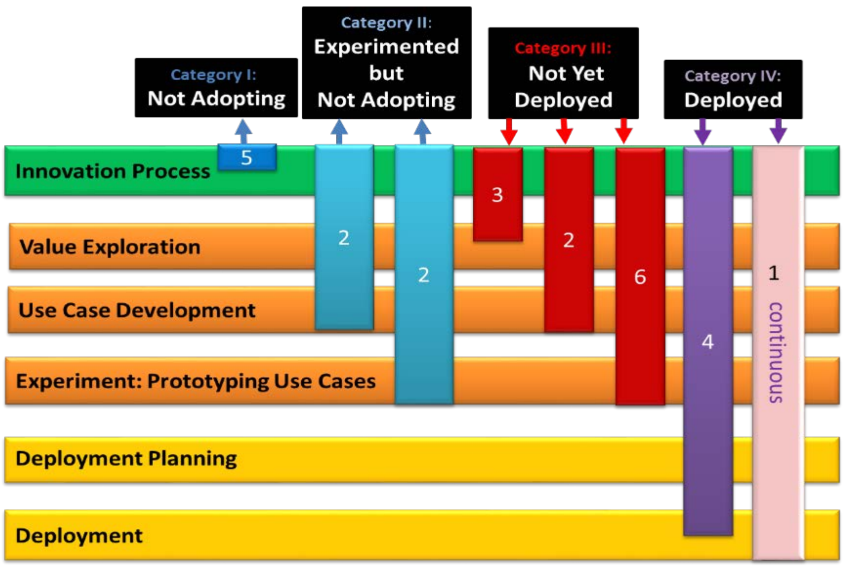

Figure 1. Adoption Status of 25 Case Enterprises

Four Big Data Adoption Categories

As shown in Figure 1, we found four adoption categories considering our cases' adoption status and processes: **Not Adopting** (20%), **Experimented but Not Adopting** (16%), **Not Yet Deployed** (44%), and **Deployed**. Of the **Deployed** category 4 enterprises (16%) had some form of deployment but only one enterprise (4%) had a full-scale deployment. The deployment rate we observed is slightly higher than Gartner's survey data. However, excluding outsourcers, a 12% deployment rate is surprisingly low as our cases were selected and pre-judged to have higher propensity and innovation capabilities to adopt.

In the **Not Adopting** category, 5 cases "engaged" with big data in some form and have some form of organizational innovation process but did not go through a full process. In the **Experimented but Not Adopting** category, 4 cases went through a formal, top-down innovation process but decided not to proceed with the big data effort (as of the present study). Their disengagement occurred in various stages of the adoption process, either in (or completing) the Value Exploration stage or Use Case Development or in the Experiment stage. The **Not Yet Deployed** category include 11 cases, also in various stages, which had intention to adopt but not yet moved to deployment. The **Deployed** category includes 5 cases that have entered the Deployment stage.

The 4 categories of adoption status are under-investigated phenomena. Previous adoption research often assumed that adoption decisions were binary: to adopt or not to adopt. We found a large percentage (44%) of cases were in the **Not Yet Deployed** category and over 50% of these had the intention to adopt but were not willing to commit or were not able to move toward to deployment. Existing studies of innovation adoption applying prominent theories such as DOI (Rogers, 2003) on the organizational level or TAM (David 1989) on the individual level only focus on Intention to Adopt, as these models assume Intention to Adopt will lead to actual adoption (**Deployment** in our study). This assumption is based on the Theory of

Planned Behavior (TPB) that assumes a strong relationship between behavioral intention and actual behavior. Our empirical study reveals that this does not apply to big data adoption. We present an alternative perspective, emerging from the data, and identify a "**Deployment Gap**" in the Big Data Adoption (BigDA, or Big², as DA means "big" in Chinese) model in Figure 2.

Big Data Adoption (Big²) model

As shown in Figure 2, our Big² model identified and integrated 11 factors that affect Intention to Adopt and subsequent Deployment.

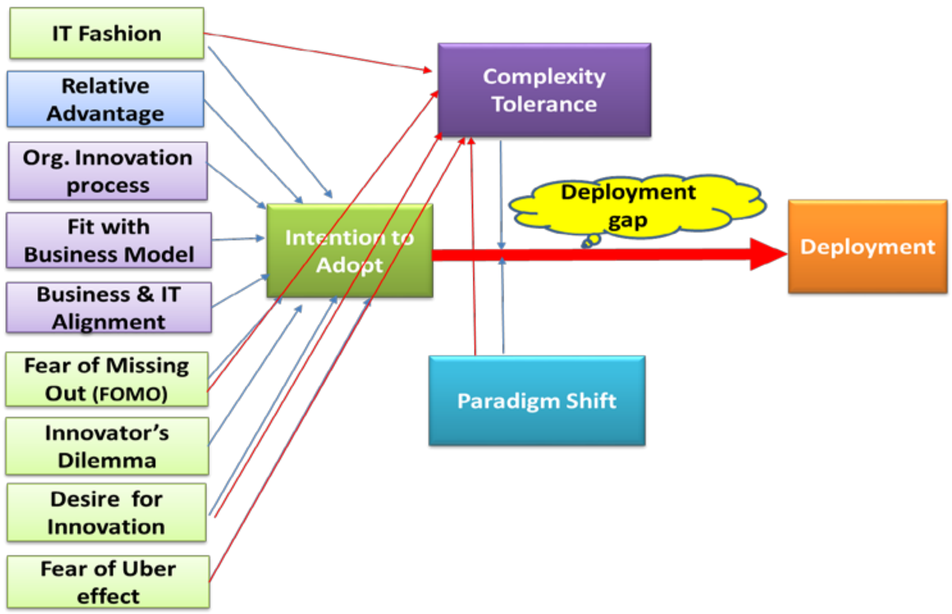

Figure 2. Big Data Adoption (Big²) Model

1. **IT Fashion**: Defined as "hype": a transitory collective belief that an information technology is new, efficient, and at the forefront of practice (Fichman 2004a; Wang 2010). Fashion setters create hype around selected innovations and promote them as 'must-deploy or fail' opportunities in an effort to influence the adoption of the innovation (Abrahamson, 1996; Kieser 1997).

2. **Relative Advantage**: Defined as the degree to which the innovation is *perceived* as a better idea compared to the one it supersedes (Rogers 2003). This is one of the most studied factors in adoption, originating from the Diffusion of Innovation (DOI) theory, mainly focusing on the advantages and disadvantages of the technology. This factor is categorized in the Technology cluster in the TOE.

3. **Organizational Innovation Process**: The organizational process of introducing new emerging IT innovations: top-down or bottom up; formal or informal. This can be categorized in the organization cluster of the TOE framework. This is different from *innovation process* that refers to the complete innovation lifecycle—the practices, procedures, and activities that take ideas and opportunities to concepts, then to development and implementation, and eventually to deployment and operation.

4. **Fit with Business Model**: The compatibility of the IT innovation with the existing business model. This can be categorized in the organizational cluster of the TOE framework.

5. **Business-IT Alignment**: Aligning information systems capabilities with business goals, including a) alignment of system infrastructure, business architecture and business model; b) strategic, organizational, political, cultural and social alignment between IT and business. (Chen et al. 2005, Chen et al. 2010). This can be categorized in the organizational cluster using the TOE framework.

6. **Fear of Missing Out (FOMO)**: the fear of missing a significant market opportunity or profitable investment or innovations which competitors are seeking. We borrow this term from psychology which is "a pervasive apprehension that others might be having rewarding experiences from which one is

absent" (Przybylski et al. 2013). This social angst is characterized by "a desire to stay continually connected with what others are doing". This new factor is related to a combination of environmental factors, e.g., competitive pressure, and psychological factors, e.g. social contagion theory, previously studied in diffusion research (Fishman 2004). Social contagion theory posits that people engage in a process of social learning by examining the actions of peers (e.g., mimetic forces arising from the tendency to imitate peers perceived to be successful under conditions of uncertainty) (DiMaggio and Powell 1983).

7. **Innovator's Dilemma**: the reluctance to adopt new innovations that impact existing mature business models. Innovator's 'dilemma' comes from the idea that businesses or organizations will reject innovations based on the fact that customers cannot currently use them, thus allowing such ideas with great potential to go to waste (Christensen 2011).
8. **Desire for Innovation**: An enterprise desired innovation to fend off competition or create new markets. Desire is the motivation for action and the appetite for a given object of attention. It can be stimulated by external forces such as advertisement.
9. **Fear of Uber Effect**: the fear of disruptive business models from others in one's market space. Uber is an on-demand crowdsourced ride sharing service. The Uber business model has had a disruptive effect (on taxi services) that was unprecedented.
10. **Paradigm Shift**: a change in the basic assumptions or fundamental practices or paradigms in a discipline. This is an extreme form of 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.
11. **Complexity Tolerance**: the extent to which an enterprise can tolerate the complexity in the technology and in its implementation process. In our study, complexity does not have an absolute value. Each enterprise has a different capability to "tolerate" complexity and the tolerance is affected by rational and non-rational factors in our model. Each enterprise has a capacity and strategies to "tolerate" complexity. An analogy would be women wearing painful high-heel shoes to be fashionable. Each woman is equipped with different capacity to wear high heels and the different degree of need for fashion will determine to what extent a women would tolerate the pain caused by high heels.

Factors 1, 6 to 9 are non-rational factors specific to big data adoption that are not captured in the TOE framework. Factors 10 (Paradigm Shifts) and 11 (Complexity Tolerance) are new variables that our model reveals to impact an enterprise's decision to move from Intention-to-adopt to deployment, which we call the "Deployment Gap". The relationships among the factors are shown in Figure 2. Paradigm Shifts contribute to increased complexity of the adoption decision, raising the level of complexity to be tolerated. The psychological factors IT Fashion, Fear of Missing Out, Desire for Innovation and Fear of Uber Effect affect Complexity Tolerance which mediates the effect on Deployment Gap. Figure 3 summarize the major factors that impact on the Intention to adopt in each of the four adoption categories and the mediating factors between Intention-to-Adopt and Deployment, answering the "why" questions. As shown, each of the factors has different effects on Intention to Adopt in different adoption categories.

Effective Complexity Tolerance Strategies

There are two complexity tolerance strategies observed from our case enterprises: complexity reduction and complexity isolation:

a) **Complexity Reduction Strategies** that we observed include: 1) Outsourcing: "Complexity tolerance/reduction strategy becomes a purchase decision"; 2) Hire new big data (analytics), experienced leadership: "We are not experiencing shortage of talents. We will pay to get whoever we need"; 3) Architectural approach: Clear architecture (modularization) for divide and conquer: "Managing complexity (or complexity reduction) is the competitive advantage and #1 differentiator of 21st century business"; 4) Business-IT alignment-SOA: architecture approach: "Our prior successful SOA implementation has greatly reduced the complexity of big data deployment decision"; 5) Knowledge-based approach for technology selection and orchestration (Chen et al. 2015); 6) Systematic, continuous absorption of complexity via Innovation Process; and 7) Center of Excellence.

b) **Complexity Isolation Strategies** that we observed include: 1) New big data (analytics) division; and 2) Entirely new company: "Innovation dilemma: hard time to grow vs. political forces; we separated the innovation to a different unit".

		Not Adopting (5)	Experimented and not adopting (4)	Not Yet Deployed (11)	Deployed (5)
1	IT Fashion	N	N	+	N
2	Relative Advantage	x	x		
3	Organizational Innovation process	x	+	+	+
4	Fit with Business Model	x	+	+	+
5	Business & IT alignment	x	x	+	+
6	Fear of Missing Out (FOMO)		+	+	+
7	Innovator's Dilemma	N	N	N	+
8	Desire for Innovation		+	+	+
9	Fear of Uber effect			+	
10	Paradigm Shift		N	N	N
11	Complexity Tolerance		X	x	+

X: "lack of"; N "negative", + "positive" impact

Figure 3. Adoption factors in each of the four adoption categories

The Mediating Effects of Paradigm Shift and Complexity Tolerance

From the findings above, the mediating effects of Complexity Tolerance and Paradigm Shifts are clear. Our results shed light on the factors on the path from Intention to Adopt to Deployment (the Deployment Gap), which was previously under-investigated in the adoption literature. Our case enterprises, even with the best Intention to Adopt, took various courses of action: 1) Experimented but Not Adopting, 2) Not Deployed, or 3) Deployed. If they cannot "absorb" the paradigm shifts and "tolerate" the complexity associated with the implementing big data, they are unable to move to deployment. Our cases who moved to deployment all employed at least one complexity reduction or isolation strategy, as discussed above. More importantly, Complexity Tolerance is contingent on multiple factors: 1) organizational capabilities to endure complexity, 2) Paradigm Shifts, and 3) the four psychological factors. Every organization has a different level of complexity tolerance. Paradigm Shifts increase the complexity of big data technology and thus reduce Complexity Tolerance. When a Paradigm Shift was not absorbed, many of our case enterprises were unable to move to deployment. As shown in Figure 2, the 4 psychological factors—IT fashion, FOMO, Desire for Innovation, Fear of Uber Effect—have direct impacts on Complexity Tolerance. They raise the threshold of "pain" felt with complexity, thus *increasing* an organization's appetite for complexity tolerance. With the psychological factors, especially Desire for Innovation and Fear of Uber Effect, enterprises are able to "tolerate" higher complexity than they inherently would.

The Limbo Stage

We observed a "Limbo Stage" in the adoption process for some of our cases in the Not-Yet-Deployed category. They signaled intention to adopt but remained in the experimental stage for years, unable to move from Intention-to-Adopt to Deployment but simultaneously not wanting to disengage from big data adoption. We identified 5 reasons, from the interplay of the 4 psychological factors and 2 mediating variables. They stay in this stage to gain legitimacy through engagement with IT (Wang 2010) or to obtain knowledge for absorbing Paradigm Shifts or to find Complexity Tolerance strategies, mostly due to the

Desire for Innovation and for Fear of Missing out or Fear of Uber Effect. Staying in the Limbo Stage is costly in many ways—indecision and inaction creates uncertainty, anxiety, and consumes resources.

IT fashion exhibits different effects on different adoption categories

Unlike other factors in the model which show effects in one direction (either N or +), IT has negative effects on the Not Adopting and the Experimented but Not Adopting categories, positive effects on the Not Yet Deployed category and no effect on the Deployed categories. In the Deployed category, enterprises actually had slightly negative attitudes towards IT fashion: they believed that overhype of big data actually "de-inflationized" important issues and turned away companies that might benefit from big data. However, these companies believed that they evaluated the technology objectively, uninfluenced by hype. We explained the enterprises' need for gaining legitimacy via engagement with IT fashion in the Not Yet Deployed category. However, for those Not Adopting, IT Fashion was viewed negatively, and a "turnoff" for some, adversely affecting the Intention to Adopt. For those Experimented and not Adopting, IT fashion created high expectations (high performance, easy and quick implementation etc.) that led the companies to be "unprepared" for the complexity and paradigm shifts.

Discussion

Our "Deployment Gap" (and Limbo Stage) bears a resemblance to the "Assimilation Gap" (Fichman & Kemerer 1999) in innovation diffusion research. The assimilation gap refers to the difference between cumulative acquisition of technology and cumulative deployment of technology. Note that deployment in innovation research is defined by widespread usage in the acquiring firms, while in our model, deployment is the action of putting the big data project into production. The technologies studied at the time (prior to 2000)—relational DBMSs, 4GLs and CASE tools—could be purchased in a single package. However, the factors causing Assimilation Gap intersect with the factors that we found associated with the Limbo Stage, such as problems with absorptive capability (Cohen and Levinthal 1990), signaling problems (Attewell 1992), knowledge barriers (Attewell 1992) and anticipation of "increasing return" (Arthur 1996).

Paradigm Shift, as we have defined it, is an extreme form of knowledge barrier. The Signaling problem was a hallmark of IT Fashion. As our study scope does not include a diffusion aspect, we did not find a real options perspective (Fichman 2004b) based on absorptive capacity theory or increasing returns to adoption. Nevertheless, when we connect the study on Limbo Stage and Assimilation Gap, it forms a richer picture of the effect of IT fashion and Paradigm Shift across adoption and diffusion processes.

Conclusion

Our exploratory multiple case study of 25 enterprises using a grounded theory method yielded rich results that extend beyond previous theories. The present study clarifies the 5V characteristics of big data and illuminates the unique value discovery process, 4 adoption categories and 11 adoption factors and their interaction effects in big data adoption. Most studies treat big data adoption decision as a binary proposition. Our study found that there are four adoption categories: Not adopting, Experimented but Not Adopting, Not Yet Deployed, Deployed. The resulting Big^2 model integrates rational and non-rational factors (some already studied in the existing literature and some new) to provide a holistic understanding of big data adoption in enterprises. We also detailed the "why" in each adoption category, offering a paradigm shift and complexity tolerance perspective.

Contrary to the prevalent assumption that Intention to Adopt inevitably leads to Deployment, we discovered a "Deployment Gap". With the best Intention-to-Adopt, our case enterprises, for different reasons, stayed in the "Limbo Stage" of Not Yet Deployed or moved to deployment or simply decided not to adopt after experimentation. The mediating effects of Complexity Tolerance and Paradigm Shift were present. The Paradigm Shift, which may not be fully recognized by enterprises, is a magnitudinal knowledge barrier to deployment. With an enterprise's inherent capabilities to "tolerate" complexity as a starting base, the Complexity Tolerance construct models the cumulative effects of organizational psychological factors and complexity resulting from the Paradigm Shift to explain how the "Deployment Gap" can be closed. This is different from previous studies where complexity is simply a technological characteristic of innovation. If an enterprise cannot totally "absorb" the paradigm shifts as well as sufficiently "tolerate" the complexity associate with the design and implementation of big data, it is unable to move to deployment. Our case enterprises who moved to deployment all found at least one complexity reduction or isolation strategy identified in our study. Relative Advantage, studied extensively in existing literature, is shown to be a necessary but not sufficient condition to close the “Gap”. As big data adoption is an enterprise level decision, Fit with Business Model, Business-IT alignment and strong Organizational Innovation Process are shown to increase Complexity Tolerance and help to close the “Gap.”

In addition, four psychological—IT Fashion, FOMO, Desire for Innovation, Fear of Uber Effect—contributed positively to Complexity Tolerance for enterprises to remain in the Limbo Stage. These four factors reflect a cut-throat competitive environment. Big data represents an opportunity for innovation: someone can become the Uber of an industry, seemly overnight, which devours the industry.

Is big data for everyone? Probably not. Our cases in the Not Adopting categories (experimented or not) answered why. If one lacks the organizational capability (e.g., Business-IT alignment, strong Organizational Innovation Process), or has the Innovator’s Dilemma and avoids risk, or lacks Desire for Innovation, Fear of Uber effect or FOMO, it will be difficult to tolerate the complexity to consider deployment, and frankly the success of deployment is uncertain. IT fashion in these categories showed negative influences due to overselling and creating unrealistic expectations. But for those who have strong desires or fears, our cases in the Deployed category serve as examples of “how to” and the 9 best practices of complexity tolerance strategies uncovered in our study can be of use.

Our Big^2 model shed lights on how big data adoption is different from previous hype such as ERP or SOA. They are similar in that they are all hyped Enterprise IT innovations and involve paradigm shifts. The main difference is that ERP and SOA both solved *existing* problems: ERP solved system integration issues and SOA solved interoperability issues that most companies faced. ERP and SOA can be adequately prototyped and the results (Relative Advantage) can be demonstrated (e.g., trialability and observability (Rogers 2003) making their adoption decision, based on Relative Advantage, straightforward. On the other hand, big data is not solving a specific existing “problem” hence it is described by non-adopters as “a hammer looking for nails.” Few truly *need* big data but some *desire* it. Can big data enable better decision making or customer intelligence? Can big data give an enterprise a new competitive edge or make it the new Uber? It *potentially* can, but it depends.

As shown in our study, the value of big data is difficult to quantify and demonstrate *a priori* while the adoption consequences are indeterministic. To help the practice, this calls for research on: 1) methods for modeling the circumstances in which big data can and can’t make a big difference; 2) methods for gauging the dynamic “value” of big data and return on “innovation,” not the traditional ROI (Hoffman 2010); 3) innovative use cases for different industries: perhaps this can be crowdsourced for open innovation.

The shift of big data adoption foci from technology to business model and from problem solving to innovation also calls for new research and invites drastic redesign of CS/IS curricula. New research is needed to help enterprises in *dealing with* (not just analyzing) paradigm shifts and complexity in big data adoption and, most importantly, to support *design* for innovation. We distinguish here between “design” and “problem solving” as design is essential to innovation and value creation (Chen and Vargo 2010). An example of “design” would be sewing squares of fabric onto a quilt while an example of “problem solving” would be putting a jig-saw puzzle together; one involves creativity and the other finding a specific solution creation. Treating design as a problem-solving activity adversely limits creativity. “Futuring” techniques (Chen et al. 2012) that encourage creativity for value exploration, experimentation and design are needed. We foresee a critical challenge to revolutionize CS/IS curricula to cultivate innovative, creative “designers” rather than simply “problem-solvers” to address paradigm shifts for big data systems.

The managerial implications from our study are significant and broad. Our study contributes lessons in each of the big data adoption categories to help enterprises navigate through the uncharted waters. More research is needed to extend the paradigm shift and complexity tolerance perspectives for guiding enterprises on whether and how to innovate with big data.

References

Abrahamson, E. 1991. “Managerial Fads and Fashions: The Diffusion and Rejection of Innovations,” *Academy of Management Review* (16:3), pp. 586-612.

Abrahamson, E. 1996. “Management Fashion,” *Academy of Management Review* (21:1), pp. 254-285.

Agarwal, R., and Dhar, V. 2014. “Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research Information Systems Research,” *Information System Research*, 25(3), pp. 443–448.

Attewell, P. 1992. “Technology diffusion and organizational learning: the case of business computing”, *Organization Science* (3:1), pp. 1–19.

Anderson, C. 2008. “The End of Theory: The Data Deluge Makes the Scientific Method Obsolete,” *Wired*, http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory

Arthur, W.B. 1996. "Increasing Returns and the New World of Business," Harvard Business Review (74:4), pp. 101-109.

Baskerville, R., Bunker, D., Olaisen, J., Pries-Heje, J., Larsen, T. J., Swanson, B.E. 2014. “Diffusion and Innovation Theory: Past, Present, and Future Contributions to Academia and Practice” in *Creating Value for All Through IT*. IFIP Advances in Information and Communication Technology Volume 429, pp. 295-300.

Basole, R. C., Seussb, D. C., and Rouse, W. B. 2013. “IT Innovation Adoption by Enterprises: Knowledge Discovery through Text Analytics,” *Decision Support Systems* (54:2), pp. 1044–1054.

Bass, L., Clements, P., and Kazman, R. 2012. *Software Architecture in Practice (3rd ed.)*. Pearson.

Beal, V. 2014. “Big Data,” *Webopedia*, http://www.webopedia.com/TERM/B/big_data.html

Benbasat, I., Goldstein, D., and Mead M. 1987. “The Case Research Strategy in Studies of Information Systems,” *MIS Quarterly* (11), pp. 369-386.

Benders, J. and van Veen, K. 2001. "What's in a Fashion? Interpretative Viability and Management Fashions," *Organization* (8)1, pp. 33-53.

Boyd, D. and Crawford K, 2012. “Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon,” *Information, Communication & Society* (15: 5), pp. 662-679.

Brewer, E. 2013. "CAP Twelve Years Later: How the "Rules" Have Changed", *Computer*, (45:2), pp. 23-29.

Chen, H-M., Kazman, R., and Perry, O. 2010. “From Software Architecture Analysis to Service Engineering: An Empirical Study of Methodology Development for Enterprise SOA Implementation.” *IEEE Transaction on Services Computing* (3: 2), pp. 145-160.

Chen, H-M., Kazman, R., Haziyev, S., and Hrytsay, O. 2015. “Big Data System Development: An Embedded Case Study with a Global Outsourcing Firm,” *Proceedings of 1st International Workshop on Big Data Software Engineering at ICSE 37*.

Chen, H-M., Kazman, R, and Garg, A. 2005. “BITAM: An engineering-principled method for managing misalignments between business and IT architectures,” *Journal of Science of Computer Programming* (57:1), pp. 5-26.

Chen, H-M., and Kazman, R. 2012. “Architecting for Ultra Large Scale Green IS,” *Proceedings of 1st International Workshop on Green and Sustainable Software at ICSE 34*.

Chen, H-M., and Vargo, S.L. 2010. “Service -Oriented Challenges for Design Science: Charting the “E”-volution,” *Pacific Asian Journal of the Association of Information Systems* (2:1), pp. 1-15.

Colombus, L. 2014. “2014: The Year Big Data Adoption Goes Mainstream in the Enterprise”, *Forbes*. http://www.forbes.com/sites/louiscolumbus/2014/01/12/2014-the-year-big-data-adoption-goes-mainstream-in-the-enterprise.

Christensen, C. 2011. *The Innovator's Dilemma: The Revolutionary Book That Will Change the Way You Do Business*, Harvard Business.

Corbin, J., and Strauss, A. 2008. *Basics of Qualitative Research*. 3 ed., Sage.

Crawford K. 2013. “The Hidden Biases in Big Data,” *Harvard Business Review*, https://hbr.org/2013/04/the-hidden-biases-in-big-data.

Damsgaard, J., and Lyytinen, K. 2001. "The Role of Intermediating Institutions in The Diffusion of Electronic Data Interchange (EDI): How Industry Associations Intervened in Denmark, Finland, and Hong Kong," *Information Society* (17:3), pp. 195-211.

Damanpour, F., and Schneider, M. 2006. “Phases 0f the Adoption 0f Innovation in Organizations: Effects of Environment, Organization and Top Managers,” *British Journal of Management* (17: 3), pp. 215-236.

Darke, P., Shanks, G., and Broadbent, M. 1998. "Successfully Completing Case Study Research: Combining Rigour, Relevance and Pragmatism", *Information Systems Journal* (8), pp. 273-289.

Davenport, T.H., Barth, P., and Bean, R. 2012. "How Big data is Different," *Harvard Business Review* (90: 10), pp. 78–83.

David, F. 1989. "Perceived Usefulness, Perceived Ease 0f Use, and User Acceptance of Information Technology", *MIS Quarterly* (13: 3), pp. 319-340.

Dodson, S. 2014. "Big Data, Big Hype?" *Wired*, http://www.wired.com/2014/04/big-data-big-hype/

Elmasri R., and Navathe S. 2010. *Fundamentals of Database Systems*, 6th edition, Addison Wesley.

Fichman, R., and Kemerer, C. 1999. "The Illusory Diffusion of Innovation: An Examination of Assimilation Gaps," *Information Systems Research* (10:3), pp. 255-275.

Fichman, R. 2004a. "Going beyond the Dominant Paradigm for Information Technology Innovation Research: Emerging Concepts and Methods," *Journal of the Association for Information Systems (*5), pp. 314–355.

Fichman, R. 2004b. "Real Options and IT Platform Adoption: Implications for Theory and Practice," *Information Systems Research* (15:2), pp. 132-154.

Frambach, R., and Schillewaert, N. 2002. "Organizational Innovation Adoption: A Multi-Level Framework of Determinants and Opportunities For Future Research," *Journal of Business Research*, (55: 2), pp. 163-176.

Fichman, R., and Kemerer, C.F. 1999. "The Illusory Diffusion of Innovation: An Examination of Assimilation Gaps," *Information Systems Research* (10: 3), pp. 255–275

Gartner. 2014a. "Gartner's 2014 Hype Cycle for Emerging Technologies Maps the Journey to Digital Business," http://www.gartner.com/newsroom/id/2819918.

Gartner. 2014b. "Survey Analysis: Big Data Investment Grows but Deployments Remain Scarce in 2014," http://www.gartner.com/document/2841519.

Glaser, B. and Strauss, A. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*, Aldine.

Glaser, B. 1992. *Basics of Grounded Theory Analysis: Emergence vs. Forcing*, Sociology Press.

Hoffman, D.L. and Fodor, M. 2010. "Can You Measure the ROI of Your Social Media Marketing?," *MIT Sloan Management Review* (52:1), pp. 41-49.

IBM, 2012. "The Four V's of Big Data," *IBM Data & Analytics Hub*. http://www.ibmbigdatahub.com/infographic/four-vs-big-data.

Infochimps, 2015. "CIOs and Big Data: What Your IT Team Wants You to Know," http://www.infochimps.com/resources/report-cios-big-data-what-your-it-team-wants-you-to-know-6.

Jacobs A. 2009. "The Pathologies of Big Data", *Communications of the ACM* (52: 8), pp.36-44.

Kelly, J. 2014. "Big Data Vendor Revenue and Market Forecast 2013-2017", http://wikibon.org/wiki/v/Big_Data_Vendor_Revenue_and_Market_Forecast_2013-2017.

Kieser, A. 1997. "Rhetoric and Myth in Management Fashion", *Organization* (4:1), pp. 49-74.

Lyytinen, K., and Damsgaard, J. 2001. "What's Wrong with the Diffusion of Innovation Theory? The case of a complex and networked technology," In *Diffusing Software Product and Process Innovations* (IFIP Volume 59), pp. 173-190.

Marcus, G. 2013. "Steamrolled by Big Data," *The New Yorker*, http://www.newyorker.com/tech/elements/steamrolled-by-big-data.

Marz, N. and Warren, J. 2013. *Big Data: Principles and Best Practices of Scalable Realtime Data Systems*, Manning Publications.

Mathiasen, L. 2002. "Collaborative Practice Research," *Information Technology & People* (14:4), pp. 321-345.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D., and Barton, D. 2012. "Big Data: The Management Revolution," *Harvard Bus Rev (90:10)*, pp. 61-67.

McKinsey, 2011. "Big Data: The Next Frontier for Innovation, Competition, and Productivity," http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation.

Orlando, B., Renzi, A., Vagnani, G., and Volpe, L. 2013. "Determinants Of Innovation Adoption: A Literature Review and Future Avenues of Research," XXV Convegno annuale di Sinergie L'innovazione per la competitività delle imprese.

Oliveira, T., and Martins, M. F. 2011. “Literature Review of Information Technology Adoption Models at Firm Level,” *The Electronic Journal Information Systems Evaluation* (14:1), pp. 110-121.

Przybylski, A., Murayama, K., DeHaan, C., and Gladwell, V. 2013. "Motivational, Emotional, and Behavioral Correlates of Fear of Missing out," *Computers in Human Behavior* (29:4), pp. 1841–1848.

Rogers, E. 2003. *Diffusion of innovations*, 5th ed., Free Press.

Sadalage, P.J., and Fowler M. 2013. *NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence*, Pearson.

Sbaraini, A., Carter, S., Evans, R. and Blinkhorn, A. 2011. “How to Do a Grounded Theory Study: A Worked Example of a Study of Dental Practices,” *BMC Medical Research Methodology* (11:128), http://www.biomedcentral.com/1471-2288/11/128.

Swanson, E. 1994. “Information Systems Innovation among Organizations,” *Management Science* (40:9), pp. 1069-1092.

Tornatzky, L. and Fleischer, M. 1990. *The Process of Technology Innovation*, Lexington, MA: Lexington Books.

Wang, P. 2010. "Chasing the Hottest IT: Effects of Information Technology Fashion on Organizations," *MIS Quarterly (34:1)*, pp. 63-85.

Winshuttle.com, 2014. “Big Data and the History of Information Storage,” http://www.winshuttle.com/big-data-timeline/.

Yin, R. 2014. *Case Study Research Design and Methods*, 5th edition, Sage Publications.

Zmud, R. 1982. “Diffusion of Modern Software Practices: Influence of Centralization and Formalization,” *Management Science* (28: 12), pp. 1421-1431.