

**REVOLUTIONARY RESEARCH STRATEGIES FOR E-BUSINESS:
A PHILOSOPHY OF SCIENCE VIEW IN THE AGE OF THE INTERNET**

Robert J. Kauffman

W.P. Carey Chair in Information Systems
W.P. Carey School of Business, Arizona State University
and Professor and Director, MIS Research Center
Carlson School of Management, University of Minnesota (on leave)
Email: rkauffman@csom.umn.edu

Charles A. Wood

Assistant Professor of Management
Mendoza School of Business, University of Notre Dame
Email: cwood1@nd.edu

Last revised: November 30, 2006

ABSTRACT

Just as the Internet has changed the way many businesses conduct business, so can it change the way that academic researchers design and execute research in e-business management. We present a series of revolutionary research strategies that employ six new data-collecting methodologies that can be used in conjunction with Internet technology. Data-collecting agents can gather very large amounts of data from the World Wide Web in a fraction of the time and the cost that it takes to gather data using traditional research methodologies. Online experiments, online judgment tasks, and online surveys expand the reach of the researcher and reduce the cost when compared to traditional experiments, judgment tasks, and surveys. Because of the vast amount of data available online, research designs such as massive quasi-experiments can be conducted that allow the researcher to find subjects without taking them out of their own environment who meet some pre-determined requirements, or some data that match a set of experimental or empirical test conditions. Finally, log files can be used to track a person's movements and actions through a Web site. We investigate these relatively new tools from a philosophy of science perspective. Using Runkel and McGrath's [110] three-horned dilemma model for traditional research methodologies as a basis, we develop a framework that illustrates the strengths and weaknesses of these new tools relative to the research design flexibility that they permit. We find that they enable advances in empirical research that would be otherwise difficult to implement when using traditional research methods.

Keywords: Data collection, electronic commerce, empirical research, IT, philosophy of science, research methodologies, online surveys, software agents.

Acknowledgments. The authors thank Mark Bergen, Jungpil Hahn, Wolfgang Jank, Dongwon Lee, Sourav Ray, Galit Shmueli, Rahul Telang, and Bin Wang, who encouraged us to consider the range of applications of these ideas. We also appreciated the reactions to these ideas offered by doctoral students Pallab Sanyal, Ajay Kumar and Ryan Sougstad— and especially Gordon Davis in the context of the IDSc 8511 Doctoral Seminar— at the University of Minnesota. Chris Dellarocas, Mark Nissen, Roberto Evaristo, Terence Ow, Paul Tallon, Ron Weber, and Weidong Xia also offered helpful advice on earlier drafts.

1. INTRODUCTION

The United States Census Bureau defines *e-business* as any process that a business organization conducts over an information technology (IT) or computer-mediated network, such as the Internet [97]. This includes e-commerce transactions, reputation system postings, and other activities that businesses may conduct over the Web. There has been tremendous growth in e-business transactions over the last several years, despite the burst of the dotcom bubble, and this growth trend is likely to continue. For example, comScore (www.comscore.com), an online business tracking company, reports that business-to-consumer (B2C) e-commerce spending increased 20.1% from US\$67.2 billion in January-June 2005 to US\$80 billion during the same period in 2006 [83]. The U.S. total for 2006 is expected to be \$170 billion, with about \$102 billion of that coming in the non-travel category. Travel e-commerce will account for about 45% of the total, with \$30.3 billion in January-June 2005 growing to \$34.7 billion in January-June 2006, a 14.7% growth rate. Non-travel spending is growing faster at 24.6%. Online auctions are included in the B2C surge. For example, in 2005, eBay, the premier online auction retailer with a dominant market share of the online auction market share, reported record net revenues of US\$4.55 billion, up 39.1% from 2004 [112]. Reports for business-to-business (B2B) e-commerce sales are even more impressive. The most recent actual statistics suggest that the U.S. economy's B2B e-commerce sales reached about \$1.8 trillion in 2004, representing about 20% of the total of all B2B sales of \$9.1 trillion [82].

For researchers, the growth over the last ten years in e-business has produced a wealth of data that few might have imagined would become available to study online shopping. The research now includes books [50], pharmaceuticals [123], digital music [124], and videos [38], as well as other kinds of consumer behaviors related to brand preferences [29], price dispersion [14,15], distribution channels [125], peer-to-peer (P2P) networks [4,52], cross-national differences [91], digital music sampling [127], shopbot performance [118,119], and consumer trust [128]. For example, most online auctions are public and viewable by researchers and the research has produced studies on bidding behaviors [3,13,115,116], and fraud [55]. Today, online portals and Internet retailers, such as Yahoo (www.yahoo.com), Google (www.google.com), Amazon (www.amazon.com) and Barnes and Noble.com (www.bn.com), list all kinds of search information and prices online, allowing researchers to track price changes and competition [49,129], book sales and the influence of online reviews [28], and evidence of portal use [80]. The same is true in the digital music and digital entertainment arena, where the availability of harvestable data is also rapidly growing [22]. Meanwhile, the UseNet newsgroup lists have messages that number in the trillions. Moreover, online comments are available on everything from reactions to transactions with bricks and mortar business to reactions to online auction sales to growth projections from stock analysts. Finally, the growth in use of radio frequency identification (RFID) technologies will give rise to

even more data, as the data collection capabilities move in the mobile organizational systems and technology arena [36].

While much has been written about the benefits that the Internet will have on the ease of commerce, very little has been written on the benefits that the Internet will have on research data collection and research design, and the academic facilitation of a more advanced managerial understanding of key e-business phenomenon. Some notable exceptions are Kauffman et al. [67], Wood and Ow [134] and Allen et al. [1]. In many different disciplines, researchers have been faced with problems involving inadequate sample size, unrealistic representation of constructs, faulty time constraints, and unacceptable cost constraints. In this article, we discuss revolutionary strategies for research design and data collection for e-commerce and e-business-related research in a way that reflects a maturing awareness of the new capabilities of IT and the World Wide Web. We do this from a *philosophy of science* perspective, so that it is possible to see how current advances have the capacity to conceptually change the way we think about conducting research. We also consider some of the newer approaches to data collection and hypothesis testing where there are new opportunities and emerging techniques that make it possible to obtain more subjects in a population sample. This in turn, will support more valid and statistically significant findings and new depths of managerial insight. Some of the new research tools that we will discuss also reflect somewhat different approaches, making substitution of new research designs possible. Data collection on the Internet has the potential to provide virtually equivalent data collection capabilities that can generate more subjects with lower costs, fewer strict assumptions, greater realism, and less contamination of research subjects' response capabilities.

We develop theory-based arguments and illustrations that are intended to answer the following research questions:

How can data collection be enhanced through the use of Internet-based tools? What are its strengths and weaknesses?

What research designs are facilitated when using the new tools that were previously infeasible with traditional research methodologies? Where will the new tools probably not have an impact?

How can we understand the nature of the underlying changes in research from the multiple points of view that are present in the philosophy of science? What predictions will they enable us to make that affect use of the new tools?

We answer these questions by examining the use of new tools for data collection in the context of existing literature, and the research designs that are particularly applicable in the presence of the new data collection methodologies. We develop a framework that illustrates the strengths and weaknesses of these new tools from a philosophy of science perspective. The articles that we will examine employ these data-

collection techniques with respect to *theory generalization*, *theory building*, *theory verification*, and *hybrid theory construction* to address specific research design challenges in e-business research. Overall, our conclusions point to the revolutionary nature of the advances that are occurring in research design and data collection, and how to think about them to develop the most effective designs for research on a variety of e-business management problems.

2. LITERATURE

Before we discuss how more new Internet-based data collection techniques change our capabilities to develop effective research designs, we first consider the basis for this kind of inquiry in philosophy of science terms, and with respect to the current measurement models that are typically used in information systems (IS) and e-commerce-related research. Our assessment is based on existing literature and it points out a new way for researchers to think about how to answer key research questions that can improve the process of e-business management.

2.1. Inquiry Systems and Empirical Research

C. West Churchman [31] the leading commentator on the practice and process of management science research in the last generation, defines *inquiry* as an activity that uses observational data to produce knowledge. The approach to developing new knowledge in the Churchmanian view allows for the possibility of behavioral adjustments to be made by researchers, when the circumstances of scientific inquiry necessitate change or evolve. This occurs, for example, when new technologies allow the investigation of emerging research questions and innovative research designs to be undertaken, when they would be difficult or impossible without the new technology. Churchman describes five different *inquiring systems*— lenses or means for understanding what is true in different research inquiries— that can be used for research: the Leibnizian, Lockean, Kantian, Hegelian and Singerian-Churchmanian inquiring systems. The first four inquiring systems are named for different philosophers of science, each of whom viewed knowledge acquisition through a particular inquiring system lens. The fifth, Singer, was Churchman's doctoral advisor, and someone whom he had a great personal impact on Churchman's view of inquiring systems.

Table 1. An Overview of Churchman's Five Inquiring Systems

INQUIRING SYSTEM	APPROACH	MAIN FEATURES
Leibnizian	Theory generalization	Information derived from axioms or models; formal, symbolic, mathematical representations; theorem and proof; best models developed; well-defined problems
Lockean	Theory building	Experimental; raw data, empirical approach used; inductive methods applied to well-defined problems
Kantian	Theory verification / falsification	Multi-model, synthetic systems; interaction between theory and data; blend Leibnizian and Lockean inquiry; assume alternative theoretical explanations; applicable to moderately ill-structured problems
Hegelian	Hybrid theory construction	Anithetical representations of a problem, opposing theories; focus is on underlying theoretical assumptions; applicable to wickedly ill-structured problems
Singerian-Churchmanian	Meta-research	Flexible approach, emphasizing continual learning and adaptation through feedback; involves conversion of ill-structured problems into structured ones, and vice versa; other inquiring systems nested within it

Note: The contents of this table were adapted from Mason and Mitroff [92, pp. 480-483].

Mason and Mitroff [92] have suggested that Churchman's models of inquiry are useful as a basis for creating evidence in the context of management IS research. We believe that this is also true for research designs and data collection involving, e-commerce, e-business management and the Internet. In this section, we describe each of the *inquiring systems* in terms of their key characteristics and approaches in research, as a basis for understanding what opportunities Internet-based tools offer for research via the literature.

Theory Generalization Research. *Leibnizian inquiry* involves theorizing first through deduction and then collecting data to support an appropriate theory. This is done by forming hypotheses based upon widely-accepted theories and models, or from theories and models that are used in similar situations. This way, researchers are able to confirm existing theory and extend accepted theory into new areas. We will refer to research that examines data from the point of view of theory with a Leibnizian perspective as *theory generalization research*, since research articles of this type tend to generalize theory to show its applicability in other areas.

Theory Building Research. *Lockean inquiry* involves data collection or observation first, and then development of theories through induction that describe the observations. From a research perspective, Lockean analysis involves developing theories when there is a wide consensus about an unexplained relationship. It also involves testing the limits of existing theoretical models based upon observations, especially when these observations do not fit within an existing theoretical framework. The existing theoretical models, then, are adjusted to accommodate the new observations. When contrasted with Leibnizian inquiry, Lockean inquiry is used to understand a situation where there is no argument about the observa-

tions of the phenomenon, but little in the way of explanatory theory. Leibnizian inquiry, in contrast, is used when theory exists to explain this phenomenon or similar phenomena. If opposing theory already exists, then the Lockean approach cannot be used, which points out the need for more in-depth induction. We will refer to research that examines data using Lockean inquiry as *theory building research*.

Theory Verification and Falsification Research. Building on our prior comments, we next come to *Kantian inquiry*. This approach is used to delve more deeply into a research phenomenon of interest through *induction*. In contrast with Lockean inquiry— which is also inductive by its nature— Kantian inquiry is not concerned with preserving existing theory. Instead, it seeks to rigorously test existing theory and either verify or falsify it. Then, if needed, the next step is to develop new theory to explain the new observations. Kantian analysis is mainly concerned with maintaining objectivity in the face of existing explanatory theory. It tends to be employed when researchers think an existing theory proves itself to be inadequate based on inconsistent research, or is fatally flawed in its explanatory power based upon new observations. We call this kind of inquiry *theory verification and falsification research*.

Hybrid Theory Construction Research. *Hegelian inquiry* involves conflicts and clashes between theoretical interpretations of similar phenomena in research. The Hegelian approach is to combine two or more conflicting theories to arrive at an explanatory combination of theories. This process encourages the researcher to assess conflicting theories through *deduction* to determine the range of their appropriateness. Thereafter, it is appropriate for the researcher to collect data to see whether it supports the new *hybrid theory* – or, as Poole and van de Ven [105] have suggested, to identify the degree of the paradox in the explanations offered by alternative theories in the presence of the same data. Hegelian research, then, concentrates on bringing conflicting theories together as it examines limitations that exist for each theory, and when each might be appropriately used. We will refer to Hegelian inquiry as *hybrid theory construction*.

Meta-Research. Singerian-Churchmanian inquiry is a meta-inquiry approach. It directs the researcher toward the selection of the proper inquiring system that will support advances in the acquisition of knowledge related to some problem or phenomenon of interest. Since Singerian-Churchmanian research does not analyze data, we will not concentrate on this mode of inquiry. However, the reader should recognize that this article might be considered as an instance of *Singerian-Churchmanian inquiry*, since it discusses how other inquiring systems can be flexibly implemented in e-business research, and what synthesis is possible to shed light on how e-business research insights are acquired with new data-collecting tools and research designs.

All these inquiring systems eventually involve some sort of data collection. But at their heart lay either the process of induction or process of deduction. *Induction* stresses the importance of data and ob-

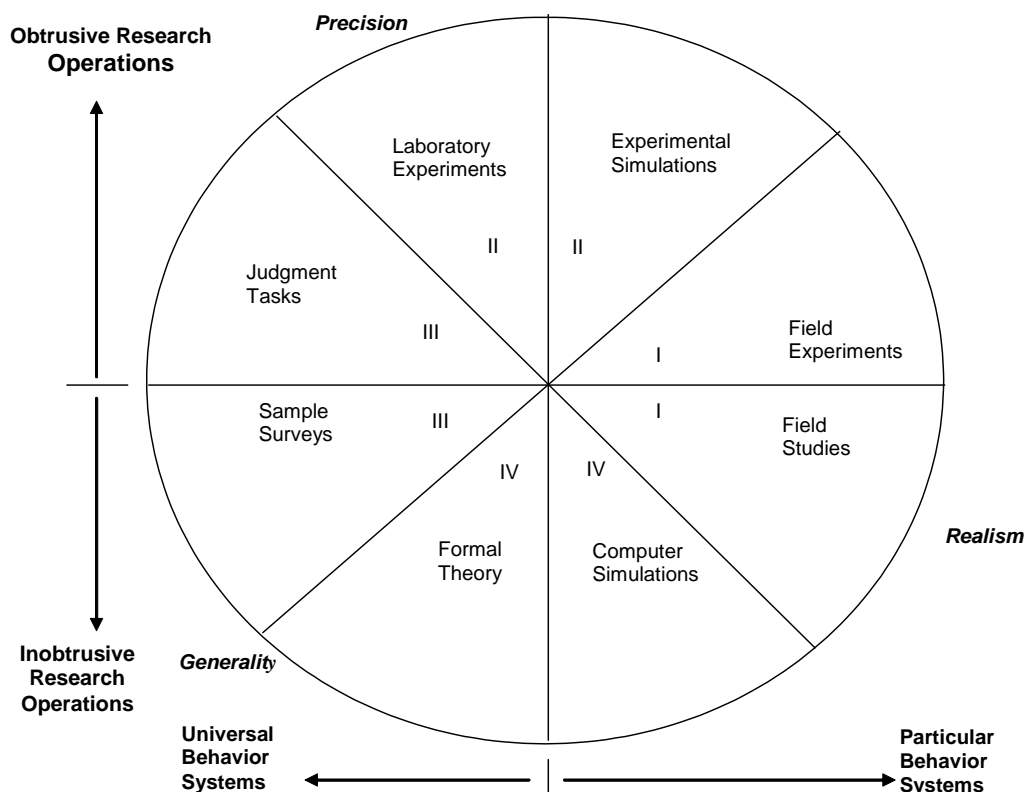
servation. Together, they enable the researcher to develop new theories. However, one concern with induction is that, when a small sample is used, its results cannot be generalized to an entire population. As a result, theories developed through induction are suspect for their lack of validity across other samples in the relevant population. *Deduction* conversely uses logic to arrive at a theory and then tests that theory using data and analysis. To support deductive efforts, data collection also is required. With the use of appropriate statistical analysis tools, the larger the data set, the less likely the researcher will draw misleading conclusions from the analysis of the data [21,74]. In the past, Churchman [31] has stressed that researchers should pursue ways to automate the collection of data, or some means to assist human researchers with data collection, so that they will be freed up to concentrate on research design and the interpretation of the research results.

The challenge with deduction is that the initial theory must be reasonably "fully formed," because otherwise empirical researchers will exclude essential aspects of the phenomenon under study [99]. Induction answers this problem by allowing the examination of data to find patterns or relationships that may exist. Induction requires even more data than deduction though. When induction is used, researchers will often split their data sets into two or more portions. One portion of a data set is used to develop an initial reading on what relationships may exist. Then deduction is used to develop theories as to why these relationships exist. Thereafter, other portions of the data set can be used to statistically verify the new theoretical relationships that are asserted. As the reader might guess, each data set must be large enough so that the statistically significant relationships can be observed. In addition, the size of each data set is especially important when using modern statistical analysis techniques. The tension that exists, however, is that we have rarely— and usually only with some difficulty and cost— been able to develop data sets large enough to afford a researcher the necessary degrees of freedom to make this process of testing, deduction, and retesting easy to accomplish.

2.2. A Philosophy of Science Assessment of Alternative Research Methods

Data collection plays an intrinsic role in the execution of the inductive and deductive research methodologies in empirical research. It not only involves retrieving data, but also interpreting and coding data. When introducing new ways to collect data, it is appropriate to review and assess existing research methodologies to see how they compare. McGrath [94], who helps us to put this comparison in a philosophy of science assessment perspective, discussed how alternative research methodologies may have recognizable strengths in one area but may be flawed in other areas. Figure 1 shows the *three-horned dilemma* of precision, generality and realism attributed to Runkel and McGrath [110].

Figure 1. Runkel and McGrath's Three-Horned Dilemma for Research Methodologies



Note: I -- Natural settings; II -- Contrived settings; III -- Settings not altered; IV -- No data collection.

The three horns of Runkel and McGrath's framework are *realism*, *generality*, and *precision*. Because of limitations of the traditional ways to collect data, no research methodology can be general, realistic and precise all at the same time. Runkel and McGrath describe these research methodologies as *dilemmatic*, because of the choices that relate to the data collection performed in each research approach to establish truth and consistency with a proposed theory.

Consider the following research methodologies in terms of the manner in which they enable the researcher to collect data:

Formal theory, analytical modeling, and computer simulations. These methods require no data collection. Instead, they generate their own data and results based on theories they assert, and, as a result, can be used to test ways of thinking about relationships between and among constructs.

Field studies. These involve primary data collection and case studies, for which a researcher collects data from a research site or multiple sites.

Experiments involve testing the effects of some stimulus against some control. Runkel and McGrath mention three types of experiments:

- *Laboratory experiments* allow researchers to examine situations stripped of their complicating environmental context.
- *Field experiments* are often undertaken in the context of workgroups, software development teams or the firm. Therefore, they are more realistic, but they are also less precise than laboratory experiments, since the latter attempts to control every aspect of an environment to eliminate "noise" in the theorized relationships.
- *Experimental simulations* try to mimic the content of the real world without actually placing the subject in the context of the real world, also to avoid complications in the research design (for example, reducing the requirements to control for certain environmental conditions).

Judgment tasks are types of experiments that involve interviews and verbal protocols to be used when subjects possess data (such as mental processes, inside information on historical events, etc.) that are not readily available for retrieval by the researcher.

Sample surveys are used to collect data about specific characteristics of a sample population to inductively test a theory.

Data-collecting agents on the Internet permit the pursuit of new types of field study research. They can change the dilemmatic nature of the applicable research methodologies, as we shall point out in the next section of this chapter. They also can affect the choices a researcher makes when choosing which among a set of alternative research methodologies will be most effective for the creation of new knowledge in a given research context.

3. DATA COLLECTION APPROACHES FOR E-BUSINESS RESEARCH

We have stressed how data collection is important to theoretical development, and how different traditional research methodologies used to collect data have strengths and weaknesses that can cause some research to be dilemmatic when investigating certain topics. Now we turn to a more in-depth discussion of new ways to collect data and the research design capabilities that are made possible through Internet technology. (See the Appendix for guidance on agent development.) We frame our discussion in terms of how these new methodologies can resolve some of the dilemmas of traditional research that Runkel and McGrath [110] pointed out. However, these new research methodologies also have strengths and weaknesses. We conclude this section with an extension of Runkel and McGrath [110] to address this issue.

3.1. New Approaches Made Possible by the Internet

We next consider five new approaches that are now possible because of the wide availability of Internet technology. They include two new data collection techniques— data-collecting agents and Web log files— and three new research approaches that extend existing methodologies— online experiments, online

judgment tasks, and online surveys. Our overall argument is that the new capabilities that Internet technology offers requires a shift in our thinking about what is possible in research design terms, as well as some reconsideration of how we go about making choices about the kinds of inquiring systems that empirical analysis involving the Internet can support for e-commerce and e-business management. We next discuss the five new approaches in greater detail, focusing on what they allow the researcher to do in e-business management research that is different from before. We begin by discussing the new ways to collect data.

Data-collecting agents. *Data-collecting agents* are software tools that are implemented when researchers want to collect data by examining Web pages that are available to the public on the Internet [73]. Data-collecting agents are sometimes referred to as *shopbots*, or just *bots* for short, and *spiders* or *spyzers*. They tend to be heavily automated, allowing direct downloading from the Web into a database, spreadsheet, or data file, based on the specifications that a researcher designs into the tool for collection of specific kinds of data (e.g., product prices and discounts, shipping costs, number of participants in an electronic auction, and so on). Also in this category, however, are the tools used by researchers who implement existing and publicly-available software agents. These include the major search engines (e.g., Google, www.google.com), shopbots (e.g., MySimon, www.mysimon.com) and searches for products or comments within a single company.

The latter capability, available through several online retailers, such as Amazon (www.amazon.com) or eBay (www.ebay.com), utilizes Internet browsers such as Internet Explorer or Netscape to discover a smaller dataset that can be manually entered into a database, spreadsheet, or data file to support a research investigation. It turns out that much information is available publicly and is readily accessed online [134]. This includes auction transactions, and comments and evaluations about products, people and companies. Also available is data related to prices charged by B2C vendors, and on the available bundles of information goods. Kauffman et al. [73] point out how data-collecting agents can gather copious amounts of information for relatively little cost compared to traditional data collection methodologies. Such large data sets allow specialized data analysis techniques to be applied with real world data. The methods include large-scale time-series or panel data econometrics, structural equation modeling, paired-observations duration modeling, social network analysis, data mining and pattern recognition, and so on, that would be challenging to implement with traditional experiments due to sample size limitations.

Web log files. Companies use Web log files to track the navigational, transaction-making, and decision behavior of human users throughout their systems. These log files usually contain more information than would ever be made available publicly on the Web. For example, log file traces of user navigation on a Web site can include sites visited before and after reaching the company's site, movements (such as

mouse movement and clicks or selections) while on the site, and indications of the patterns of use of the organizational hierarchy of information that is available on a Web site. In practice, such data may reach into the tens of thousands or even millions of records. Occasionally, vendors and firms make these log files available to researchers, giving them data analysis opportunities that heretofore have not been available. As with data-collecting agents, the acquisition of such large datasets allows analysis of realistic situations in e-commerce and e-business that would be difficult to implement with most traditional assisted data collection techniques, and would be infeasible for manual data collection. We now turn to a discussion of three other research approaches that have resulted from recent applications of Internet technology.

Online experiments. Researchers also now have the capability to conduct online experiments over the Internet to set up different situations in which subjects can respond. Even though the researcher is bound to lose some control over the subjects that participate in these tests (e.g., eBay auction participants, customers of an electronic grocer's Web site, or users of an online search engine) when compared to traditional experiments, access to subjects is much greater, and data acquisition costs are considerably lower.

Online judgment tasks. Online judgment tasks can be used to examine motivations for actions on the part of human subjects. Often, judgment tasks are done within context that most observers might interpret to be experimental research settings (e.g., involving decision making and discrimination, interpretation of the conditions that are present in some context, reactions to various stimuli, materials or other contextual elements, etc.). By their nature, these judgment tasks often take on an experimental quality. We note that typical judgment tasks often will be presented with the realistic context being stripped out of the data collection, so that the researcher can more clearly examine some relationship of interest. Thus, online judgment tasks are typically used inside hypothetical environments, hypothetical decision making settings, or contrived situations where the researcher is looking for specific kinds of reactions and responses. In addition, these judgment tasks cannot be used to directly examine actions, but only to infer actions from examination of task results and the questioning of the participants in the study.

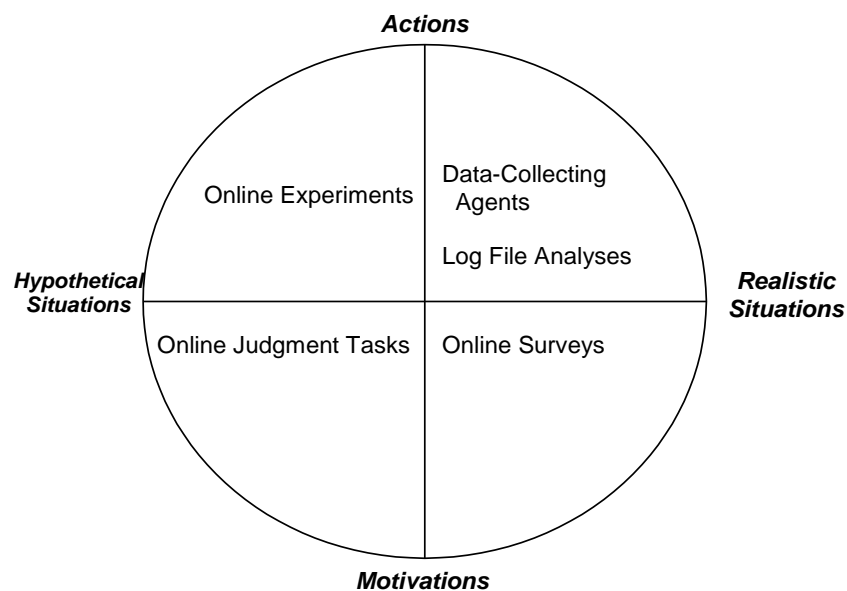
Online surveys. Online surveys, including surveys done by email and surveys that are presented using data collection tools embedded in a Web site, can be used to question potential respondents about their actions and motivations. Online surveys usually can be done more quickly and cheaply when compared to traditional surveys, and often with a broader reach for subjects and participants. However, like online judgment tasks, online surveys cannot be used to directly examine actions. One can only infer actions from survey responses, since the data collected are subjective rather than objective. There have been many studies that compare responses in online surveys or email surveys to traditional surveys (e.g., [30,95] or even between different online implementations of the same survey, such as with an embedded

or attached survey [41]. Most of these studies show little or no significant change in response, thus allowing researchers to pick a survey instrument on the basis of ease of implementation rather than on the basis of data quality.

3.2. Assessment of the Five New Approaches to Research in E-Business

In keeping with Runkel and McGrath's three-horned dilemma model, we contend that Internet technology has altered the process of data collection. For instance, the low cost to implement an Internet agent or a Web log allows observations that span across different companies (e.g., [25,33,34,73]), making these field studies more generalizable. We concentrate on how the strengths and weaknesses of online data collection methodologies compare to each other and to traditional research methodologies. Figure 2 shows how these research methodologies compare to each other, once again using the constructs that we presented related to Figure 1.

Figure 2. Trade-Offs in E-Business Research Methodologies



We note that:

data-collecting agents and log files are best suited for e-business research in realistic situations where data are available;

online experiments are useful for studying human actions inside hypothetical environments or when realistic data are unavailable; and,

online surveys and judgment task experiments support the examination of the motivations behind human actions in realistic situations, and the motivation for human actions inside hypothetical environments.

Table 1 shows the strengths and weaknesses of each research approach.

Table 1. Strengths and Weaknesses of E-Business Research Approaches

RESEARCH APPROACHES	STRENGTHS	WEAKNESSES
<i>New Ways to Collect Data</i>		
Data-Collecting Agents	Inexpensive to get large transaction data sets Realistic, actual transactions, observations Non-intrusive and easily repeatable Multiple simultaneous data sources available Does not require vendor cooperation	Can be burdensome to host computer Data must be present online; not all types of data are available Can only infer motivations
Log File Analyses	More data types (buyer identity, etc.) available Extremely large data sets, non-intrusive process Realistic, actual transactions/observations	Usually single data source Can only infer motivations Requires cooperation from vendor
<i>New Extensions to Existing Research Methodologies</i>		
Online Experiments	Far reaching experiments Cheap compared to traditional experiments Can test hypothetical environments	Technology may bias results Controls may not be as strong Responses are not made in context
Online Judgment Tasks	Can be used to examine motivations Can be used with hypothetical environments Inexpensive compared to the set up of traditional judgment tasks	Can be biasing, corrupting, unrealistic due to IT involved Actions only inferred from questions (self reporting bias, memory bias) Task training may be problematic
Online Surveys	Can be used to examine motivations Less expensive than traditional surveys Faster than traditional surveys	Actions only inferred from responses Realism of results can be questionable

Just as in Runkel and McGrath's original framework, a researcher may use a research methodology that is not necessarily suggested by the above framework. For example, while a decision maker's or a consumer's motivations usually cannot be directly examined with data-collecting agents or log file analysis, they can be inferred in situations where surveys or judgment tasks are impractical or where the participant may wish to hide her actions or motivations, such as in the case of fraud or opportunism (e.g., [71,133]).

Runkel and McGrath point out that it is important for the researcher to understand and acknowledge the research dilemma and its implications when justifying the approach that is to be employed. Weaknesses notwithstanding, the strengths of these relatively new techniques allow examination of research questions that were previously impossible or impractical to answer due to costs, contamination of subjects' responses and so on.

3.3. Other Methods to Extend E-Business Management Research Capabilities

There are likely to be new ways to conduct research that will be very useful, given the nature of elec-

tronically-available data. This section describes two other new approaches, in particular, that we believe are likely to have an especially salient impact in e-business research.

Massive quasi-experiments. Quasi-experiments in IS and e-commerce research aim to take advantage of naturally-occurring conditions in the real world to capture data that enable a researcher to distinguish between outcomes associated with different levels of influence, access or use of IT. In the case of the Internet, there are many natural settings that permit quasi-experimental research designs to be developed for the study of human decision making, business processes, and organizational performance outcomes in the presence of technology. The sources of data for quasi-experimental research designs include log files or customized data from a data-collecting agent. They permit researchers to restrict the data they collect to support quasi-experimental research designs.

In our experience (e.g., [71,73]), when so much data becomes available, it is relatively easy to find individuals who meet multiple control and stimulus conditions. As a result, we think of quasi-experimental research designs in the context of very, very large data sets as *massive quasi-experimental designs*. It may be the case that it is necessary to sift through a very large data set (for example, a year's worth of daily prices on 1,000 products among 20 online booksellers, or two year's worth of eBay auction data for all of the bidders on auctions involving U.S. cent coins) to find just those observers that meet a set of predefined experimental conditions (for example, when price change reactions are apparent, or when there is some need to identify whether fraudulent bidding might be occurring).

Moreover, these individuals can remain in their own context and need not be taken to a laboratory setting for an experimental test. This is important given that other research process observers (e.g., [77,111]) have pointed out that subjects do not behave as they normally would when taken out of their own context. However, such realism necessitates acceptance of somewhat less control over the subjects. A countervailing consideration is that collection of extremely large data sets may permit the collection of data that represents substantially all of the behavior or firm actions that the researcher wishes to observe or track.

Use of Artificial Intelligence and Expert Systems Approaches. There has been much research on the use of artificial intelligence to solve business problems, especially when using Web-based data (e.g., [98,120]). The problem with Web data is that it is often in text form and there are thousands, hundreds of thousands, or millions of records. Using an expert system or computer-aided processing approach to deal with this text, with some type of pattern-matching algorithm, allows the researcher to identify items with less difficulty, lower cost, and with a higher rate of precision, than by reading each item and determining item characteristics. For example, Table 2 contains 20 records of actual data retrieved from Dell Computer (www.dell.com) laptop auctions.

Table 2. Contents of a Representative *Description* Column for Dell Product Listing Table

<i>DESCRIPTION</i>
DELL WIN2K LATITUDE C600 750MHz 128MB TEST DO NOT BID
DELL Optiplex GX110/MT 933MHz TEST DO NOT BID
DELL WIN2K LATITUDE C600 750MHz 128MB TEST DO NOT BID
DELL Optiplex GX110/MT 933MHz TEST DO NOT BID
DELL WIN2K LATITUDE C600 750MHz 128MB TEST DO NOT BID
DELL Optiplex GX110/MT 933MHz TEST DO NOT BID
TEST ONLY DELL (NO O/S) Latitude C600 750MHz 256MB 20GB 24X 56K 10/100 NIC
TEST DO NOT BID DELL WIN2K LATITUDE C600 750MHz 128MB 20GB 24X 56K
TEST DO NOT BID DELL WIN2K Latitude C600 750MHz 128MB 20GB 24X 56K
TEST DO NOT BID DELL (NO O/S) Latitude C600 850MHz 512MB 20GB 24X 56K
TEST DO NOT BID DELL (NO O/S) Latitude CPXJ 750GT 750MHz 256MB 20GB 24X No Modem
TEST DO NOT BID DELL (NO O/S) Latitude C600 750MHz 320MB 20GB 24X 56K
TEST DO NOT BID DELL (NO O/S) Latitude C600 750MHz 256MB 20GB 24X 56K
DELL (No O/S) Dimension L1000 1000MHz 128MB 20GB 48X
DELL (No O/S) Optiplex GX110/L 866MHz 128MB 15GB DVD
DELL (No O/S) Optiplex GX110/MT 800MHz 128MB 20GB 48X
DELL (No O/S) Optiplex GX150/SDT 933MHz 256MB 10GB 24X
DELL (No O/S) Optiplex GX150/SMT 1000MHz 256MB 20GB ZIP DVD
DELL (No O/S) Dimension 4300 1500MHz 256MB 20GB 32X
DELL (No O/S) Optiplex GX110/L 866MHz 192MB 15GB DVD
Source: www.dell.com . Accessed on September 14, 2004.

If the data set consisted of only 20 records, as in Table 2, a researcher could easily identify each item being sold by hand. However, if a software agent retrieves 10,000 records for such computers that are for sale *each day* for several months *from several different computer sites*, a researcher would be hard-pressed to identify or match similar items for sale without some sort of automated assistance. This problem, however, is readily solved by using simple pattern-matching algorithms. For example, the pattern recognition algorithm can be written in Visual Basic for Applications (VBA) that comes bundled with various Microsoft products, or using a stored procedure that most databases implement. To illustrate, consider the following VBA code that can be used to process SQL commands inside a Microsoft Access table called *DellComputer*. The code removes test records, and then finds the chip speed and stores it inside the *mhz* column in the database:

```

' Remove test data
DoCmd. RunSQL "Delete from Dell Computer WHERE Instr(Description), 'TEST' > 0"

' Process Mhz
For x = 700 to 2000
    MySQL = "UPDATE Dell Computer SET mhz = " & x & " WHERE mhz is null AND "
    MySQL = MySQL + " Instr(Description, '" & x & "MHz') > 0"
    DoCmd. RunSQL MySQL
Next

```

Such processing can be somewhat complicated when the decisions about the data become more complex. This is likely to require an expert system that discriminates among the text-based data using rules, and determines the order of execution of those rules. In addition, an artificial intelligence capability may be appropriate, so that the program is able to actually change its own decision processes without any human intervention, based upon new information it receives. When finding simple patterns in the data, the code required to identify items can be quite simple, as illustrated above. Automation often reduces the human error that can accompany many hours of scanning through data. Incorporation of such tools is part of the research design when implementing powerful data collection techniques.

4. UNDERSTANDING E-BUSINESS RESEARCH THROUGH THE INQUIRY SYSTEMS

The grid shown in Table 3 crosses the first four of Churchman's five types of inquiry systems in research with the five research approaches. This will give the reader examples of how a methodology was implemented in the context of various problem and topic areas in e-business research. As shown by the distribution of the literature that we have sorted according to the dimensions in Table 2, certain approaches appear to be predisposed to support research investigations within specific types of theory development. Based upon this literature review, it also appears, for example, that data-collecting agents seem to be the most used methodology for data collection in e-business research. This is understandable. Data-collecting agents are inexpensive to run, can gather large amounts of data, and do not require the permission of the hosting Web site. We will next survey the literature according to our framework's classification to discuss theory development and how researchers have designed research and used new data-collecting tools for empirical e-business management research.

A word is in order on the methodology that we used to select the papers and to place them in the framework. We identified candidate articles from the leading journals and conferences in the IS field, and then expanded our coverage to include the fields of Economics and Marketing. The articles included in this chapter are representative, and not truly exhaustive of all of the available research on e-commerce during the past ten years involving Internet-based data collection. To have pursued this latter task would have been a very great effort, indeed, and one that would not have served our ultimate purpose of illustrating the revolutionary new approaches and empirical advances that are seen in e-commerce research. Moreover, many of the articles that we identified, as it turned out, either did not collect data in the manner that we have described, or did not collect data at all. The most common instance for rejecting the inclusion of an article came when the authors used manual techniques (e.g., making queries on the Web via a search engine, manually clicking through Web pages, or obtaining secondary data in a traditional manner).

Table 3. Inquiring Systems, Data Collection Approaches and E-Business Research

RESEARCH APPROACH	THEORY GENERALIZATION	THEORY BUILDING	THEORY VERIFICATION / FALSIFICATION	HYBRID THEORY CONSTRUCTION
<i>New Ways to Collect Data</i>				
Data-Collecting Agents	Ba and Pavlou [5] Bajari and Hortacsu [6,7] Bergen et al. [19,20] Brinkmann and Seifert [23] Dewan and Hsu [40] Granados et al. [54] Livingston [86]	Ariely and Simonson [3] Asvanund et al. [4] Bapna et al. [11,12] Bergen et al. [18] Brynjolfsson et al. [24] Dans [37], Eaton [45] Dellarocas and Wood [38] Ederington and Dewally [46] Gilkeson and Reynolds [51] Gopal et al. [52] Kauffman and Wang [68] Kauffman and Wood [71,72] Lucking-Reiley et al. [89] Lynch and Ariely [90] Massad and Tucker [93] Oh [101], Park and Kim [104] Roth and Ockenfels [109] Segev et al. [113], Standifird [121] Wang et al. [130], Wood et al. [133]	Bapna et al. [9] Bergen et al. [16,17] Brynjolfsson and Smith [25] Chevalier and Mayzlin [28] Clay et al. [32,33] Ghose et al. [49] Houser and Wooders [58] Hyde et al. [59] Jank and Shmueli [60,61] Kauffman and Lee [65,66] Levy et al. [81] Lucas and Oh [87] Lucking-Reiley [88] Melnik and Alm [96] Ow and Wood [112], Jap [62] Pan et al. [103], Smith [118] Smith and Brynjolfsson [2001] Wilcox [132]	Alpar et al. [2] Baye et al. [14] Clemons et al. [35] Dellarocas and Wood [39] Easley and Tenorio [44] Kauffman and Lee [65,66] Kauffman and Wood [70,73] Levy et al. [81] Resnick and Zeckhauser [107] Ward and Clark [131] Yamagishi [135]
Log File Analyses	Hahn and Kauffman [56] Murphy et al. [100]	Catledge and Pitkow [26] Hahn and Kauffman [57] Lee et al. 2003 [80] Tillotson et al. [126]	Brynjolfsson and Smith [25] Hahn and Kauffman [57] Lee [79]	Clay et al. 2001, 2002a, 2002b
<i>New Extensions to Existing Research Methodologies</i>				
Online Experiments	None	Bapna et al. [10] Jin and Kato [63], Kamis [64] List and Lucking-Reiley [85] Rafaeli and Roy [106]	Doong et al. [42] List and Lucking-Reiley [84] Resnick et al. [108]	None
Online Judgment Tasks	Ba and Pavlou [5]	Bapna [8]	None	None
Online Surveys	Chen and He [27], Ferl and Mill-sap [47], Lederer et al. [78]	Earp and Baumer [43] Sheehan [114], Yin [136]	Gardyn [49] Gordon and Lima-Turner [53]	None
Note: This table provides a more exhaustive list than the articles which are discussed in the body of the chapter.				

We also screened out papers which we believed did not have significant results in terms of contributions to new knowledge. Thereafter, the co-authors split the load of bucketing the various papers in the different boxes of the framework. We then discussed our separate assignments, especially checking for discrepancies. In some cases, it made sense to position a study in more than one box, because of the different research approaches that were demonstrated. When we disagreed on the assignment of a paper, we considered one another's arguments in favor of one bucketing or another, and eventually reached a consensus.

4.1. Theory Generalization Research

As shown in Table 3, online experiments are not often used to generalize existing theory in different areas in the e-business research context. The nature of experiments is to apply a stimulus and a control to a subject, and the impersonal nature of e-business often makes such application difficult. However, other ways to collect data are more apt to facilitate theory generalization in e-business research.

Data-collecting agents can be used to examine existing theory in the context of the new e-business environment. Bajari and Hortacsu [6] examine how the typical empirical economic regularities from a sample match the observed patterns in data found in other samples of online auctions. The authors also examine the exact effects of the *winner's curse* (where the auction winner overpays) in online auctions [7]. Brinkmann and Seifert [23], Dewan and Hsu [40], and Livingston [86] all find that trust increases the consumer's willingness-to-pay, as predicted by trust theory. Ba and Pavlou [5] also find a trust effect. They conduct an online judgment task for students to complete, and follow it by analyzing data that has been collected online using a data-collecting agent. Hahn and Kauffman [56] and Murphy et al. [100] apply different theories that explain the efficacy of Web design to develop hypotheses about Web site performance. In both cases, the authors show how the analysis of log files helps to shed light on optimal Web site design. More recently, Doong et al. [42], another chapter in this research volume, have explored the efficacy of different participation incentives for online group-buying auctions using an online experimental design approach with multiple treatments.

Online surveys can also be used for theory generalization. Chen and He [27] develop a theory of online technology adoption with respect to a specific retailer. Then they validate this theory using online survey responses that are evaluated via structural equation modeling. Lederer et al. [78] apply a similar technique and validate the *technology acceptance model* (TAM) for work-related tasks and Web applications. They also employ a survey with structural equation modeling to test TAM in this environment. Ferl and Millsap [47] also conduct an online survey to show how technology acceptance is dependent upon ease of use, as predicted by the TAM model.

4.2. Theory Building Research

Theory building is the area where most of the activity has concentrated in e-business research. This is understandable, since e-business is a relatively new phenomenon, and thus, it presents us with many challenges to try to understand business relationships within e-business.

Data-collecting agents have been shown to be a useful tool for building theory because of the large amounts of data that can be collected from a multitude of sources to test relationships that provide a basis for establishing new theory. Ariely and Simonson [3] show that bidders under-search, and, therefore over-pay in online auctions. They also suggest that high starting bids are likely to lead to higher prices only if competing items are not available. Bapna et al. [11,12] examine bidder types and bidder behavior in online auctions, especially bidding behavior relates to the prices paid and the timing of bidders' bids. Dans [37] discusses new business models that are made possible by Internet technology.

In addition, Dellarocas and Wood [39] evaluate factors that influence bidder participation levels in online auctions using an extensive data set collected on eBay's Web site. Wood et al. [133] examine how sellers with higher reputations behave in such a manner as to reduce the amount of positive buyer comments from transactions. Jin and Kato [63], Eaton [45], and Park and Kim [104] analyze eBay data and show how additional information can reduce the information asymmetry inherent in e-business transactions, resulting in a buyer's willingness to pay more for an item and increasing a buyer's commitment to a seller.

In a similar vein, Kauffman and Wang [68] analyzed the effects of buyer arrival rates and price plateaus in group-buying Websites. Chevalier and Mayzlin [28] evaluate the impacts of word of mouth from online book reviews on book sales. Ederington and Dewally [46], Lucking-Reiley et al. [89], Gilkeson and Reynolds [51] and Standifird [121] use data collected with data-collecting agents to examine factors that affect the final price bid in online auctions. Kauffman and Wood [71] examine the possibility of shilling through the examination of a massive data set that examines bidder behavior during concurrent auctions selling the same item. Massad and Tucker (2000) [93] compare online auctions and traditional auctions, and find, surprisingly, that online auctions lead to higher dollar values in initial bid prices and in final bid prices, when compared with physical auctions. Segev et al. [113] develop a new theoretical model for prediction of final auction price, and test this model using auction data collected with a data-collecting agent. Catledge and Pitkow [26] and Tillotson et al. [126] were among the first researchers to examine Internet browsing behavior through the use of log files.

Online experiments also are extensively used in e-business theory building, especially in hypothetical situations. In a laboratory experiment, List and Lucking-Reiley [85] show how bundling items within auctions can result in large price premiums. Bapna et al. [10] also use a laboratory experiment to analyze dif-

ferent price-setting processes in online auctions. Rafaeli and Noy [106] developed experimental results to show that interpersonal additions to online auctions which mimic face-to-face contact increase transaction amounts.

Quasi-experiments can also be useful in finding relationships that are difficult to examine in traditional environments. Kauffman and Wood [61] establish controls and stimuli for weekend and weekday buying, and buying with and without a picture of the trade item. They find that there is a *weekend effect*, and that pictures and reputation scores also affect the final price. Roth and Ockenfels [109] examine *bid sniping*, or bidding at the last possible moment, with auctions on Amazon and eBay, and find that the fixed closing time on eBay motivates more sniping than the variable closing time on Amazon. By examining judgments in an online environment, Bapna [8] is able to propose a new auction microstructure designed to eliminate sniping behavior.

Earp and Baumer [43] and Sheehan [114] use online surveys to test the privacy concerns of online shoppers. Both find some skepticism among users about providing personal information in online transactions, although Sheehan finds that education levels will mollify privacy concerns. Yin [136] surveys and tests the effects of price dispersion with selling price. Geographical differences and the norms associated with the sharing of private information also will affect the efficacy of the approaches that we have discussed. For example, European Web site operators are precluded from sharing data with e-commerce researchers due to European Union data protection laws. In the United States, in contrast, by default consumers and users of various Web sites opt in for the use of their data by default, and need to express their desire to opt out. In Europe, the opposite is true: the default is to opt out, with the result that far fewer consumers and users' data are available for study by a third party. Paul Tallon, a co-editor of this research volume, pointed out that such fundamentally different assumptions could pose a barrier to some forms of data collection – even though the relevant data will probably be there, somewhere in an inaccessible data base.

Theory building is extremely data-intensive, involving induction from observations, and then, to test the theories, empirical examination to determine if the new theories are applicable to an environment. We show here how new data collection techniques, especially data-collecting agents, Web log file analysis, and online surveys, are useful to collect the data needed for the task of theory building.

4.3. Theory Verification Research

Often, theory verification results in the questioning of some theory's validity within the context of a given environment of research interest. Because of the easy access to data that the new data-collecting tools support, theory verification has become more attractive. For example, Bapna et al. [9] used data in online auctions to reject the often-used auction theory assumption that bidders are homogeneous. They

then go on to illustrate different bidder types. Ow and Wood [102] explore the effect that winner's curse has in online auctions and find that buyer experience leads to an increase in willingness-to-pay in online auctions. This is the opposite of the findings that have obtained in traditional markets. Houser and Wooders [58] examine both online auctions and online reverse auctions, which involve contract bids where the lowest bid wins. They find that although a seller's reputation affects the final price in online auctions, a buyer's reputation does not affect price in reverse auctions. List and Lucking-Reiley [84] examine several theoretical auction models advocated in economic theory, yet find no significant difference in revenue between these different types of auctions. Similarly, Lucking-Reiley [88] finds that Dutch auctions generate approximately 30% higher revenues in email and newsgroup auctions than in traditional first-price and second-price auctions, contrary to theoretical predictions. Melnik and Alm [96] analyze eBay data to determine the effect a seller's reputation has on price, and find only a small effect, contrary to other theory on this topic. Wilcox [132] examines the timing of bids, and finds that experienced bidders bid in a way that is more consistent with theory over time than inexperienced bidders.

Using log file analysis, Brynjolfsson and Smith [25] find evidence of friction in e-commerce markets, which also is the opposite of what was expected predicted by prior armchair theory builders. Also using log files, Lee [79] examines Aucnet in Japan, a used automobile online auction, and finds that quality guarantees from a third-party auditor tend to increase willingness-to-pay in an e-market.

Resnick et al. [108] examine what happens when they take a well-established eBay seller, and then set up a new eBay seller, and observe how they interact in terms of buyers' willingness-to-pay. Their research design involves having each seller sell the exact same item. They found an 8.1% positive difference in willingness-to-pay for the reputable seller. In another quasi-experiment, Jap [62] investigates how price competition mechanisms affect buyer-supplier relationships. He finds that reverse auctions can increase the supplier's belief that the buyer will act opportunistically. Gardyn [48] conducts an online survey of 871 children, ages 8 to 15. She finds, contrary to conventional wisdom, that older girls are more active computer users than their male counterparts. A survey by Gordon and Lima-Turner [53] indicates that social contract theories may be somewhat limited in areas of e-business.

Verifying existing theory requires much evidence, especially if the verification process yields a result that is contrary to the reigning conventional thought on a topic. As such, these new research technology-based techniques are ideal for re-examining conventional theory, especially as this theory applies to e-business phenomena. Here, we show how theory verification can be accomplished using new research methodologies.

4.4. Hybrid Theory Construction

Rather than attempting to verify or validate a theory, hybrid theory construction techniques allow the researcher to examine theories that are often diametrically opposed. This permits the determination of the circumstances where one theory is suited to a situation and when another theory would be more appropriate. Often, hybrid theory construction requires larger sets of data, and so it is understandable that data-collecting agents and log files— the techniques that are best able to generate the largest datasets— are observed to be the most suitable for this type of inquiry. For example, when examining competition and tacit collusion, which are opposing actions in competitor interaction, Kauffman and Wood [70] find that market leaders often dictate the dynamics for price competition for the rest of the industry, and that the same online companies react to competitors differently as members of different industries. Resnick and Zeckhauser [107] examine reputation systems in online auctions and report that feedback was left despite incentives to free ride, that feedback is almost always positive, and that feedback does predict a seller's future performance in online auctions. Also, contrary to other literature, they find that higher reputation scores do not result in higher prices paid for an item. Easley and Tenorio [44] find that experienced bidders tend to bid in higher increments, which they call *jump bidding*, and they theorize about the cost of repeat bidding compared against the cost of jump bidding. Clay et al. [34] describe that firm reactions to competitor price changes do occur, but not as often as one would expect. Yamagishi [135] examines the effect of open versus closed auctions. Although the theory that he uses states that closed auctions will be less prone to opportunism and generate higher prices, Yamagishi still finds that a reputation system can make an open market outperform a closed auction.

Of all the theory development research we have discussed so far, hybrid theory construction can be the most data-intensive. When there are conflicting theories, with each applying to different segments of a market or population (such as competition and collusion), data are required to examine the theories to gain an accurate picture of how these they interact. The data can be used to evaluate the extent of the paradox that is created, in the words of Poole and van de Ven [105]. These authors have championed and applied this kind of thinking to build new management and organizational theories – something that the new approaches to research for e-commerce can support as well, only with as much data as is needed. Clearly, the new data collection techniques greatly simplify the task of data collection, and thus facilitate hybrid theory construction.

We now have illustrated how different types of theoretical development are accomplished with typically less time, less cost, and more ease than when doing traditional data collection. We have shown that these techniques are particularly well-suited to e-business research by demonstrating how researchers have successfully implemented them to accomplish their research objectives.

5. CONCLUSION

This research delves into the newest frontier of research design approaches and data collection techniques for e-business management. We examine these new approaches from a philosophy of science perspective, and extend Runkel and McGrath's [110] framework to illustrate different facets of these new techniques that extend beyond traditional research approaches. Leveraging Churchman's [31] insights, we also illustrate how e-business researchers implement these approaches when developing theory that investigates important e-business problems. The result of this examination is that we call attention to several research directions that can facilitate empirical advances in e-business research.

5.1. Contributions

We investigated the strengths and weaknesses of these new approaches. We pointed out, echoing the long-standing insights of others who have written on research approaches from a philosophy of science perspective, how these tools can be used in theory building, theory generalization, theory verification, and theory combination. For this reason, we claim that it is critically important to appreciate how much impact they ought to have on the way in which researchers conceptualize and go about designing their research approaches for the exploration of current problems in e-business management. We found in our exploration that these new technology-based tools are able to resolve several issues associated with traditional data collection methodologies. This appears to be occurring because of researchers' new ability through Internet-based tools to inexpensively collect megabytes of data on individuals' actions. As a result, it is possible to use these approaches to mimic the data collection outcomes of traditional methods, without having to face up to their inherent limitations.

5.2. Research Directions

With these preliminary findings in mind, we think it is worthwhile to suggest some research directions involving the new data collection methodologies that may be appropriate for future research on the spectrum of e-business management topics.

Massive quasi-experiments. With so much data available, it is relatively easy to find individuals who meet multiple control and stimulus conditions, as we have seen in our prior discussion. These individuals remain in their own context and need not be taken to a laboratory setting for an experimental test. Lave [77] and Scribner [111] describe how subjects behave differently inside laboratories, even with tasks as simple as bowling or buying milk, and show how this lack of realism can give erroneous results. Traditionally, however, quasi-experiments are difficult to conduct due to exorbitant costs involved in achieving a sample size of sufficient power, and are also difficult due to possible contamination of subject behavior. Internet data, however, can be captured without the express knowledge of the participant, leading to significant ethical issues. This helps to resolve some of the unattractive features and difficulties of tradition-

al quasi-experiments. However, this also leaves some issues that will require additional study from university and research organization-based institutional review boards for research.

Fast, cheap surveys. Often, surveys can take months to implement. Pre-testing and revising a survey instrument can be costly and laborious. With online surveys, the deployment is less expensive, the reach is greater, and the time required for the collection of responses is greatly compressed. Although we are not advocating a quick and dirty approach, we think it is important to point out that the fundamental cost-to-results quality relationship is undergoing significant and beneficial changes in favor of the researcher. The usual battery of non-response bias tests are still needed though, as are tests to assess the generalizability of the results to the population of interest.

Time-series and longitudinal research designs. Many of these new techniques, such as online surveys or data-collecting agents, can be easily re-implemented as time passes. Thus, researchers can more easily gather time-series data on individuals, as well as panel data for a cross-section of individuals over time. Such data typically are not available when traditional data collection is done, other than in the most extraordinary circumstances. Indeed, longitudinal research designs place the greatest cost and feasibility pressures on traditional data collection techniques, so it is likely that opportunities for Internet-based longitudinal research designs will dramatically change the cost-to-research quality relationship.

Theory-building and empirical data analysis. Many theories have not yet been tested extensively with realistic data. For example, Bapna et al. [9] investigate popular economic theories using online data, and find that many of them are not supported. In more recent research, Kauffman and Wood [73] investigate the appropriateness of Bertrand competition assumption in online markets, a long-standing truism in the IS literature, and find that Bertrand competition is not sufficient to explain competitive interaction.

"Hidden" phenomena. One of the main concerns with traditional data collection is that some behavior (such as opportunism) is hardly ever exhibited when these techniques are used. This is because individuals can maintain control over the data that are presented to the researcher. Thus, a non-invasive data collection methodology is required to more closely examine the actions of individuals without having those individuals change their behavior because they know that they are being examined – the well-known *Hawthorne effect*. Data-collecting agents and log-file analysis are natural tools for such investigation, coupled with new methods for data visualization [115,116]. For example, Kauffman and Wood [71] investigate online shilling by seeing which bidders make bids on items that have characteristics different than the normal bid. Then these bidders are examined for shilling-like behavior. Such a technique would be extremely difficult or impossible to duplicate using traditional data collection. In addition, Jank and Shmueli [60,61] and Shmueli et al. [117] use curve clustering and functional data analysis to reveal a variety of online auction bidder behavior phenomena through innovative statistical analysis.

We have shown how cost-effective and time-saving these new research approaches can be. A data-collecting agent can be developed by a single capable programmer in less than a month. If it is designed properly, it can collect literally hundreds of thousands of records in a single day. Surveys can be rapidly deployed with no mailing or materials costs, and with a 24-hour response window. Also, it is often possible to retest results from analysis of agents data or online survey data repeatedly or to gather additional data or to gather data that were omitted early in a study. With other research approaches, such as traditional surveys, experiments, or field studies, collecting new data may be extremely problematic. Indeed, the new ways to collect data for e-business research are so effective, the researcher can now spend the majority of her time thinking through the elements of the best research design, determining what data to collect, analyzing the available data, and figuring out how to best present the findings. As Churchman [31] seems to have foreseen, this is all in sharp contrast to traditional data collection, for which the majority of the researcher's time must be spent on the data collection effort itself.

5.3. Transforming Our Thinking for Empirical Advances in E-Business Research

The technologies that have made e-business possible simultaneously have transformed many industries, but thus far academic researchers have shown little scientific awareness of the manner in which these technologies are also transforming the underlying processes of research design and data collection— at least not in philosophy of science terms or in a way that systematically lays out how we should proceed. Many companies, such as General Electric, Dell, Microsoft and Cisco Systems, have recognized through their IT investments, how the efficiencies achieved through the Web can allow them to function with greater responsiveness and with far less expense, supporting higher levels of profitability than otherwise might be possible. At the same time, many consumers recognize how the Web allows them to more easily search for information and do electronic shopping.

So too, we argue, will our thinking about research design and data collection in empirical research for e-business management have to be transformed. The new technologies will permit us to investigate emerging areas of e-business management for far lower cost and in far less time than we probably expected even five years ago. We predict that, in time, researchers will come to think of the World Wide Web as a data source and a research context where they can research their topics of interest with perspectives on research design and methodologies for data collection that are as new as the technologies of the Internet themselves. We recognize, of course, that some variables (especially intention-based and perceptual variables, for example) will never be observable in the Web environment.

5.4. Limitations

Surely, however, the use of traditional approaches to research design and data collection will remain appropriate for many types of projects. It is unlikely, for example, that the use of Internet-based tools

will be able to replace the depth of insight that can be obtained from interviewing real people (e.g., consumers, senior managers, Web-based systems designers, or e-commerce pricing strategists about their decision making approaches) who have to deal with real world risks and outcomes in their organizational and business environments. In addition, we recognize that no real world quasi-experiment– undertaken on a limited basis or with a massive data collection approach– can approach the degree of control that a researcher can assert in a more controlled laboratory environment. Yet, even with powerful controls, it will be necessary to extrapolate the validity of the findings to the real world. Clearly, situations in which human motivation (i.e., on the part of an individual, a manager or group of managers relative to a firm ‘s strategy or performance) is the key variable are less apt to be effectively understood with Internet-based techniques for data collection.

In addition, it is appropriate to point out that some aspects of e-business management research will require joint Internet technology-based and traditional approaches to be effective in yielding meaningful managerial insights. For example, most investigations into the business value of Web-based technologies in e-business processes will continue to require data on investment levels, the timing of expenditures, and other behind-the-scenes data that capture the key drivers of organizational performance. No doubt, Internet-based data collection should be able to provide important information about the business processes themselves, how they have been configured, the output levels and performance dimensions associated with them, as well as the extent of the use of a given software application and the information associated with it [122]. Constructing the full picture will remain costly and challenging nevertheless.

In conclusion, we also challenge researchers in e-business management to give more thought to whether they will be able to justify total reliance upon the new techniques. In the future, researchers will not only need to understand the best ways to design e-business research to build, test and validate theory. They will also need to be aware of the best ways to construct and direct a data-collecting agent, how to incorporate data-collecting agents into research designs that also involve traditional data collection, and gauge the potential impact that data-collecting agents will have on what now becomes possible in e-business management research. In spite of the likely growing pains that we expect to experience as we gain confidence with these new empirical advances, we nevertheless see a bright and productive future for the application of new technologies in support of the creation of new e-business management knowledge.

REFERENCES

1. Allen, G.N., Burk, D. L., and Davis, G. B. Academic data collection in electronic environments: Defining acceptable use of Internet resources. *MIS Quarterly* 30, 3, 2006, 599-610.
2. Alpar, P., Poremski, M., and Pickerodt, S. Measuring the efficiency of Web site traffic generation. *International Journal of Electronic Commerce*, 6, 1, 2001, 53-74.

3. Ariely, D., and Simonson, I. Buying, bidding, playing, or competing? Value assessment and decision dynamics in online auctions. *Journal of Consumer Psychology* 13 (1-2), 2003, 113-123.
4. Asvanund, A., Karen, C., Krishnan, R., and Smith, M. D. An empirical analysis of network externalities in peer-to-peer music sharing networks. *Information Systems Research*, 15, 2, 2004, 155-174.
5. Ba, S., and Pavlou, P. A. Evidence of the effect of trust building technology in electronic markets: Price premiums and buyer behavior. *MIS Quarterly*, 26, 3, 2002, 243-268.
6. Bajari, P., and Hortacsu, A. Are structural estimates of auction models reasonable? Evidence from experimental data. Working paper, Department of Economics, Stanford University, Stanford, CA, 2003a.
7. Bajari, P., and Hortacsu, A. The winner's curse, reserve prices and endogenous entry: Empirical insights from eBay auctions. *Rand Journal of Economics*, 34, 2, 2003b, 329-355.
8. Bapna, R. When snipers become predators: Can mechanism design save online auctions? *Communications of the ACM*, 46, 12, 2003, 152-158.
9. Bapna, R., Goes, P., and Gupta, A. A theoretical and empirical investigation of multi-item on-line auctions. *Information Technology and Management*, 1, 1-2, 2000, 1-23.
10. Bapna, R., Goes, P., and Gupta, A. Comparative analysis of multi-item online auctions: Evidence from the laboratory. *Decision Support Systems*, 32, 2, 2001a, 135-153.
11. Bapna, R., Goes, P., and Gupta, A. Insights and analysis of online auctions. *Communications of the ACM*, 44, 11, 2001b, 42-50.
12. Bapna, R., Goes, P., and Gupta, A. Replicating online Yankee auctions to analyze auctioneers' and bidders' strategies. *Information Systems Research*, 14, 3, 2003, 244-268.
13. Bapna, R., Goes, P., Gupta, A., and Karuga, G. Predicting bidders' willingness-to-pay in online multi-unit ascending auctions: Analytical and empirical insights. Working paper, School of Business, University of Connecticut, Storrs, CT, 2004.
14. Baye, M.R., Morgan, J., and Scholten, P. Temporal price dispersion: Evidence from an online consumer electronics market. *Journal of Interactive Marketing*, 18, 4, 2004, 101-115.
15. Baylis, K., and Perloff, J. Price dispersion on the Internet: Good firms and bad firms. *Review of Industrial Organization*, 21, 3, 2002, 305-324.
16. Bergen, M., Kauffman, R. J., Kim, H. K., and Lee, D. W. Customer antagonism-driven price rigidity in Internet-based selling. Presented at the 40th Hawaii International Conference on Systems Science, Big Island, HI, IEEE Computing Society Press, Los Alamitos, CA, January 2007.
17. Bergen, M., Kauffman, R. J., and Lee, D. W. Store quality image and the rational inattention hypothesis: An empirical study of the drivers and \$9 and 9¢ endings. *INFORMS Conference on Information Systems and Technologies*, Denver, CO, October 2004.
18. Bergen, M., Kauffman, R. J., and Lee, D. W. Discovering price rigidity and price adjustment patterns on the Internet: Evidence from the e-bookselling industry. *Journal of Management Information Systems*, 22, 2, 2005, 57-90.
19. Bergen, M., Kauffman, R. J., and Lee, D. W. From quality to competition: The role of economic factors to explain digital prices. Working paper, Carlson School of Management, University of Minnesota, Minneapolis, MN, June 2006a.
20. Bergen, M., Kauffman, R. J., and Lee, D. Price points and price rigidity on the Internet: A massive quasi-experimental data mining approach. Working paper, Carlson School of Management, University of Minnesota, Minneapolis, MN, June 2006b.
21. Beutler, L. E. The view from the rear: An editorial. *Journal of Consulting and Clinical Psychology*, 64, 5, 1996, 845-847.

22. Bockstedt, J., Kauffman, R. J., and Riggins, F. J. The move to artist-led music distribution: Explaining market structure changes in the digital music market. *International Journal of Electronic Commerce*, 10, 3, 2006, 7-38.
23. Brinkmann, U., and Seifert, M. Face to interface: Zum problem der vertrauenskonstitution im Internet am beispiel von elektronischen auktionen (Face to interface: The establishment of trust in the Internet– the case of electronic auctions). *Zeitschrift für Soziologie (Journal for Sociology)*, 30, 1, 2001, 23-47.
24. Brynjolfsson, E., Hu, J., and Smith, M. Consumer surplus in the digital economy: Estimating the value of increased product variety. *Management Science*, 49, 11, 2003, 1580-1596.
25. Brynjolfsson, E., and Smith, M. Frictionless commerce? A comparison of Internet and conventional retailers. *Management Science*, 46, 4, 2000, 563-585.
26. Catledge, L. D., and Pitkow, J. E. Characterizing browsing strategies in the World-Wide Web. *Computer Networks and ISDN Systems*, 27, 6-14, 1995, 1065-1073.
27. Chen, R., and He, F. Examination of brand knowledge, perceived risk and consumers' intention to adopt an online retailer. *Total Quality Management and Business Excellence*, 14, 6, 2003, 677-694.
28. Chevalier, J. A. and Mayzlin, D. The effect of word of mouth on sales: Online book reviews. Working paper ES-28, School of Management, Yale University, New Haven, CT, 2003.
29. Chu, W., Choi, B., and Song, M.R. The role of on-line retailer brand and infomediary reputation in increasing consumer purchase intention. *International Journal of Electronic Commerce*, 9, 3, 2005, 115–127.
30. Church, A. H. Is there a method to our madness? The impact of data collection methodology on organizational survey results. *Personnel Psychology*, 54, 4, 2001, 937-969.
31. Churchman, C. W. *The Design of Inquiring Systems*. Basic Books, New York, NY, 1971.
32. Clay, K., Krishnan, R., and Wolff, E. Prices and price dispersion on the Web: Evidence from the online book industry. *Journal of Industrial Economics*, 49, 4 2001, 521–539.
33. Clay, K., Krishnan, R., Wolff, E., and Fernandes, D. Retailer strategies on the Web: Price and non-price competition in the online book industry. *Journal of Industrial Economics*, 50, 3, 2002a, 351–367.
34. Clay, K., Ouyang, S., Smith, M., and Wolff, E. Leader follower behavior? Evidence from online book markets. Working paper, Heinz School of Public Policy and Management, Carnegie Mellon University, Pittsburgh, PA, 2002b.
35. Clemons, E., Hann, I., and Hitt, L. Price dispersion and differentiation in online travel: An empirical investigation. *Management Science*, 48, 4, 2002, 534–549.
36. Curtin, J., Kauffman, R. J., and Riggins, F. J. Making the MOST out of RFID technology: A research agenda for the study of the adoption, use and impacts of RFID. *Information Technology and Management*, 8, 2, 2007, in press.
37. Dans, E. Existing business models for auctions and their adaptation to electronic markets. *Journal of Electronic Commerce Research*, 3, 2, 2002, 23-31.
38. Dellarocas, C. N., Awad, N. and Zhang, X. Using online reviews as a proxy of word-of-mouth for motion picture revenue forecasting. Working paper, Sloan School of Management, MIT, Cambridge, MA, 2004.
39. Dellarocas, C., and Wood, C. A. The sound of silence in online feedback: Estimating trading risks in the presence of reporting bias. *Management Science*, 2007, in press.
40. Dewan, S., and Hsu, V. Trust in electronic markets: Price discovery in generalist versus specialty online auctions. Working paper, University of Washington, Seattle, WA, 2001.
41. Dommeyer, C. J., and Moriarty, E. Comparing two forms of an e-mail survey: Embedded versus attached. *Journal of the Market Research Society*, 42, 1, 1999-2000, 39-52.

42. Doong, H. S., Kauffman, R. J., Lai, H. C., and Zhuang, Y. T. Empirical design of incentive mechanisms in group-buying auctions: An experimental approach. In R. J. Kauffman and P. Tallon (Eds.), *Advances in Management Information Systems*, ME Sharpe, Armonk, NY, 2007, in this volume.
43. Earp, J. B., and Baumer, D. Innovative Web use to learn about consumer behavior and online privacy. *Communications of the ACM*, 46, 4, 2003, 81-83.
44. Easley, R. F., and Tenorio, R. Jump-bidding strategies in Internet auctions. *Management Science*, 50, 10, 2004, 1407-1419.
45. Eaton, D. H. Valuing information: Evidence from guitar auctions on eBay. Working paper, Department of Economics and Finance, Murray State University, Murray, KY, 2002.
46. Ederington, L. H., and Dewally, M. A comparison of reputation, certification, warranties, and information disclosure as remedies for information asymmetries: Lessons from the online comic book market. Working paper, Price College of Business, University of Oklahoma. Norman, OK, 2003.
47. Ferl, T. E., and Millsap, L. Remote use of the University of California MELVYL Library System: An on-line survey. *Information Technology and Libraries*, 11, 3, 1992, 285-303.
48. Gardyn, R. Born to be wired. *American Demographics*, 25, 3, 2003, 14-15.
49. Ghose, A., Smith, M.D., and Telang, R. Internet exchanges for used books: An empirical analysis of product cannibalization and welfare impact. *Information Systems Research*, 17, 1, 2006, 3-19.
50. Ghose, A., and Sundararajan, A. Evaluating pricing strategy using e-commerce data: Evidence and estimation challenges. *Statistical Science*, 21, 2, 2006, 131-142.
51. Gilkeson, J. H., and Reynolds, K. Determinants of Internet Auction Success and Closing Price: An Exploratory Study. *Psychology and Marketing*, 20, 6, 2003, 537-566.
52. Gopal, R., Bhattacharjee, S., and Sanders, L. Do artists benefit from online music sharing? *Journal of Business*, 79, 2006, 1503-1533.
53. Gordon, M. E., and Lima-Turner, K. D. Consumer attitudes toward Internet advertising: A social contract perspective. *International Marketing Review*, 14, 5, 1997, 362-375.
54. Granados, N., Gupta, A., and Kauffman, R. J. Market transparency in Internet-based selling: Modeling and empirical analysis. INFORMS Main Conference, San Francisco, CA, 2005.
55. Gregg, D. G., and Scott, J. E. The role of reputation systems in reducing on-line auction fraud. *International Journal of Electronic Commerce*, 10, 3, 2006, 95-120.
56. Hahn, J., and Kauffman, R. J. Measuring and comparing the effectiveness of e-commerce Web site designs. Working paper, MIS Research Center, Carlson School of Management, University of Minnesota. Minneapolis, MN, 2003.
57. Hahn, J., and Kauffman, R. J. A methodology for business value-driven Web site evaluation. 2004 Human-Computer Interaction/Management Information Systems Workshop, Washington, DC, December 2004.
58. Houser, D., and Wooders, J. Reputation in auctions: Theory and evidence from eBay. Working paper, Eller School of Business, University of Arizona, Tucson, AZ, 2000.
59. Hyde, V., Jank, W., and Shmueli, G. Investigating concurrency in online auctions through visualization. *American Statistician*, 2006, in press.
60. Jank, W. and Shmueli, G. Profiling price dynamics in online auctions using curve clustering. Working paper, Smith School of Business, University of Maryland, College Park, MD, 2005.
61. Jank, W., and Shmueli, G. Functional data analysis in electronic commerce research. *Statistical Science*, 2006, in press.
62. Jap, S. D. An exploratory study of the introduction of online reverse auctions. *Journal of Marketing*, 67, 3, 2003, 96-107.

63. Jin, G., and Kato, A. Blind trust online: Experimental evidence from baseball cards. Working paper, Department of Economics, University of Maryland, College Park, MD, 2002.
64. Kamis, A. Search strategies in shopping engines: An experimental investigation. *International Journal of Electronic Commerce*, 11, 1, 2006, 63-84.
65. Kauffman, R. J., and Lee, D. Price rigidity on the Internet: New evidence from the online bookselling industry. In R. Agarwal, L. Kirsch and J. I. DeGross (Eds.), *Proceedings of the 25th International Conference on Information Systems*, Washington, DC, December 2004.
66. Kauffman, R. J., and Lee, D. Should we expect less price rigidity in the digital economy? Working paper, University of Minnesota, Minneapolis, MN, November 2006.
67. Kauffman, R. J., March, S. T., and Wood, C. A. Design principles for long-lived Internet agents. *International Journal of Intelligent Systems in Accounting, Finance and Management*, 9, 4, 2000, 217-236.
68. Kauffman, R. J., and Wang, B. New buyers' arrival under dynamic pricing market microstructure: The case of group-buying discounts on the Internet. *Journal of Management Information Systems*, 18, 2, 2001, 157-188.
69. Kauffman, R. J., and Wood, C. A. What makes a buyer pay more for the same item in Internet auctions? Invited paper presentation, Empirical Research in Online Auctions, Economics Cluster, INFORMS Main Conference, Miami, FL, November 2001.
70. Kauffman, R. J., and Wood, C. A. Analyzing competitive and tacitly collusive strategies in electronic marketplaces. Working paper, MIS Research Center, Carlson School of Management, University of Minnesota, Minneapolis, MN, 2003.
71. Kauffman, R. J., and Wood, C. A. The effects of shilling on final bid prices in online auctions? *Electronic Commerce and Research Applications*, 4, 1, 2005, 21-34.
72. Kauffman, R. J., and Wood, C. A. Doing their bidding: An empirical examination of factors that affect a buyer's utility in Internet auctions. *Information Technology and Management*, 7, 3, 2006, 171-190.
73. Kauffman, R. J., and Wood, C. A. Follow the leader? Strategic pricing in e-commerce, *Managerial Decision and Economics*, 2007, in press.
74. Kendall, P. C., and Flannery-Schroeder, E. C. Methodological issues in treatment research for anxiety disorders in youth. *Journal of Abnormal Child Psychology* 26, 1, 1998, 27-39.
75. Klein, and Dellarocas, C. Exception handling in agent systems. In *Proceedings of the Third International Conference on Autonomous Agents*, Seattle, WA, ACM Press, New York, NY, 1999, 62-68.
76. Lakshmanan, L.V.S., Sadri, F., and Subramanian, S.N. SchemaSQL: An extension to SQL for multidatabase interoperability. *ACM Transactions on Database Systems* 26, 4, 2001, 476-519.
77. Lave, J. *Cognition in Practice: Mind, Mathematics and Culture*, Cambridge University Press, Cambridge, MA, 1988.
78. Lederer, A. L., Maupin, D. J., Sena, M. P., and Zhuang, Y. The technology acceptance model and the World Wide Web. *Decision Support Systems*, 29, 3, 2000, 269-282.
79. Lee, H. G. Do electronic marketplaces lower the price of goods? *Communications of the ACM*, 41, 1, 1998, 73-80.
80. Lee, S., Zufryden, F., and Drèze, X. A study of consumer switching behavior across Internet portal Web sites. *International Journal of Electronic Commerce*, 7, 3, 2003, 39-63.
81. Levy, D, Lee, D., Chen, A, Kauffman, R. J. and Bergen, M. Making sense of ignoring cents: Price points and price rigidity under rational inattention. Working paper, MIS Research Center, Carlson School of Management, University of Minnesota, Minneapolis, MN, November 2006.

82. Lenard, T. M., and Britton, D. B. *The Digital Economy Factbook*, 8th Edition, The Progress and Freedom Foundation, Washington, DC, 2006.
83. Lipsman, A. comScore forecasts total e-commerce spending by consumers will reach approximately \$170 billion in 2006, comScore.com, Reston, VA, [www.comscore.com/press/ release.asp?press=959](http://www.comscore.com/press/release.asp?press=959), last accessed on November 29, 2006.
84. List, J. A., and Lucking-Reiley, D. Demand reduction in multi-unit auctions: Evidence from a sportscard field experiment. *American Economic Review*, 90, 4, 2000, 961-972.
85. List, J. A., and Lucking-Reiley, D. Bidding behavior and decision costs in field experiments. *Economic Inquiry*, 40, 44, 2002, 611-619.
86. Livingston, J. A. How valuable is a good reputation? A sample selection model of Internet auctions. Working Paper, Department of Economics, University of Maryland, College Park, MD, 2002.
87. Lucas, H. C., Jr., and Oh. W. Information technology and pricing decisions: Price adjustments in online computer markets. *MIS Quarterly*, 30, 3, 2006, 755-775.
88. Lucking-Reiley, D. Using field experiments to test equivalence between auction formats: Magic on the Internet. *American Economic Review*, 89, 5, 1999, 1063-1080.
89. Lucking-Reiley, D., Bryan, D., Prasad, N., and Reeves, D. Pennies from eBay: The determinants of price in online auctions. Working paper, Department of Economics, University of Maryland, College Park, MD, 2000.
90. Lynch, J.G., and Ariely, D. Wine Online: Search costs affect competition on price, quality and distribution. *Marketing Science*, 19, 1, 2000, 83-103.
91. Mahmood, M.A., Bagchi, K., and Ford, T.C. On-line shopping behavior: Cross-country empirical research. *International Journal of Electronic Commerce*, 9, 1, 2004, 9-30.
92. Mason, R. O., and Mitroff, I. I. A program for research on management information systems. *Management Science* 19, 5, 1973, 475-487.
93. Massad, V. J., and Tucker, J. M. Comparing bidding and pricing between in-person and online auctions. *Journal of Product and Brand Management*, 9, 5, 2000, 325-332.
94. McGrath, J. E. Dilemmatics: The study of research choices and dilemmas. In J. E. McGrath, J. Martin, and R. Kula, (Eds.), *Judgment Calls in Research*, Sage Publications Inc., Beverly Hills, CA, 1982, 69-102.
95. Mehta, R., and Sivadas, E. Comparing response rates and response content in mail versus electronic mail surveys. *Journal of the Market Research Society*, 37, 4, 1995, 429-439.
96. Melnik, M. I., and Alm, J. Does a seller's e-commerce reputation matter? Evidence from eBay auctions. *Journal of Industrial Economics*, 50, 3, 2002, 337-349.
97. Mesenbourg, T. L. Measuring electronic business: Definitions, underlying concepts, and measurement plans. Research report, Assistant Director for Economic Programs, Bureau of the Census, United States Government, Washington, DC, 2004. Available at www.census.gov/epcd/www/ebusiness.htm. Last accessed on November 29, 2006.
98. Metaxiotis, K. S., Askounis, D., and Psarras, J. Expert systems in production planning and scheduling: A state-of-the-art survey. *Journal of Intelligent Manufacturing* 13, 4, 2002, 253-260.
99. Mitroff, I. I., and Mason, R. O. Business policy and metaphysics: Some philosophical considerations. *Academy of Management Review* 7, 3, 1982, 361-371.
100. Murphy, J., Hofacker, C. F., and Bennett, M. Website-generated market-research data: Tracing the tracks left behind by visitors. *Cornell Hotel and Restaurant Administration Quarterly*, 42, 1, 2001, 82-91.
101. Oh, W. C2C Versus B2B: A comparison of the winner's curse in two types of electronic auctions. *International Journal of Electronic Commerce*, 6, 4, 2002, 115-138.

102. Ow, T., and Wood, C. A. The conflicting effect of experience on buying and selling in online auctions: An empirical examination of the effect of trust in electronic markets. In A. Bharadwaj, S. Narasimhan and R. Santhanam (Eds.), *Proceedings of the Fall 2003 Conference for Information Systems and Technology*, Atlanta, GA, October 2003.
103. Pan, X., Ratchford, B., and Shankar, V. Can price dispersion in online markets be explained by differences in e-tailer service quality? *Marketing Science*, 30, 4, 2002, 433-445.
104. Park, C., and Kim, Y. Identifying key factors affecting consumer purchase behavior in an online shopping context. *International Journal of Retail and Distribution Management*, 31, 1, 2003, 16-29.
105. Poole, M. S., and van de Ven, A. H. Using paradox to build management and organizational theories. *Academy of Management Review*, 14, 4, 1989, 562-578.
106. Rafaeli, S., and Noy, A. Online auctions, messaging, communication and social facilitation: A simulation and experimental evidence. *European Journal of Information Systems*, 11, 3, 2002, 196-207.
107. Resnick, P., and Zeckhauser, R. Trust among strangers in Internet transactions: Empirical analysis of eBay's reputation system. In M. R. Baye (Ed.), *Advances in Applied Microeconomics*, 11, Elsevier Science, Amsterdam, Netherlands, 2002, 127-158.
108. Resnick, P., Zeckhauser, R., Swanson, J., and Lockwood, K. The value of reputation on eBay: A controlled experiment. Working paper, School of Information, University of Michigan, Ann Arbor, MI, 2003.
109. Roth, A. E., and Ockenfels, A. Last-minute bidding and the rules for ending second-price auctions: Evidence from eBay and Amazon auctions on the Internet. *American Economic Review*, 92, 4, 2002, 1093-1103.
110. Runkel, P. J., and McGrath, J. E. *Research on Human Behavior: A Systematic Guide to Method*. Holt, Rinehart and Winston, New York, NY, 1972.
111. Scribner, S. Studying working intelligence. In B. Rogoff and J. Lave (Eds.), *Everyday Cognition: Its Development in Social Context*, Harvard University Press, Boston, MA, 1984.
112. Securities Exchange Commission. eBay 2005 annual report, U.S. Government, Washington, DC, 2006. Available at www.sec.gov/Archives/edgar/data/1065088/000095013406003678/f17187e10vk.htm, last accessed on November 29, 2006.
113. Segev, A., Beam, C., and Shanthikumar, J. G. Optimal design of Internet-based auctions. *Information Technology and Management*, 2, 2, 2001, 121-163.
114. Sheehan, K. B. Toward a typology of Internet users and online privacy concerns. *Information Society*, 18, 1, 2002, 21-31.
115. Shmueli, G., and Jank, W. Visualizing online auctions. *Journal of Computational and Graphical Statistics*, 14, 2, 2005, 299-319.
116. Shmueli, G., Jank, W., Aris, A., Plaisant, C., and Shneiderman, B. Exploring auction databases through interactive visualization. *Decision Support Systems*, 2006, in press.
117. Shmueli, G., Russo, R. P., and Jank, W. Modeling bid arrivals in online auctions. Working paper, Smith School of Business, University of Maryland, College Park, MD, 2004.
118. Smith, M. D. The impact of shopbots on electronic markets. *Marketing Science* 30, 4, 2002, 442-450.
119. Smith, M. D., and Brynjolfsson, E. Customer decision making at an Internet shopbot: Brand still matters. *Journal of Industrial Economics* 49, 4, 2001, 541-558.
120. Srivastava, J. and Cooley, R. Web Business Intelligence: Mining the Web for Actionable Knowledge. *INFORMS Journal on Computing* 15, 2, 2003, 191-198.
121. Standifird, S. S. Reputation and e-commerce: eBay auctions and the asymmetrical impact of positive and negative ratings. *Journal of Management*, 27, 3, 2001, 279-296.

122. Straub, D., Limayem, M., and E. Karahanna-Evaristo, E. Measuring system usage: Implications for IS theory testing. *Management Science*, 41, 8, 1995, 1328-1344.
123. Stylianou, A. C., Kumar, R. L., and Robbins, S. S. Pricing on the Internet and in conventional retail channels: A study of over-the-counter pharmaceutical products. *International Journal of Electronic Commerce*, 10, 1, 2005, 135-152.
124. Tang, F., and Lu, D. Pricing patterns in the online CD market: An empirical study. *Electronic Markets*, 11, 3, 2001, 171-185.
125. Tang, F., and Xing, X. Will the growth of multi-channel retailing diminish the pricing efficiency of the Web? *Journal of Retailing*, 77, 3, 2001, 319-333.
126. Tillotson, J., Cherry, J., and Clinton, M. Internet use through the University of Toronto Library: Demographics, destinations, and users' reactions. *Information Technology and Libraries*, 14(3), 1995, 190-198.
127. Tu, Y., and Lu, M. An experimental and analytical study of on-line digital music sampling strategies. *International Journal of Electronic Commerce*, 10, 3, 2006, 39-70.
128. Urban, G.L., Sultan, F., and Qualls, W. Placing trust at the center of your Internet strategy. *Sloan Management Review*, 24, 1, 2000, 39-48.
129. Walter, Z., Gupta, A., and Su, B. The sources of online dispersion across product types: An integrative view of on-line search costs and price premiums. *International Journal of Electronic Commerce*, 11, 1, 2006, 37-62.
130. Wang, S., Jank, W., and Shmueli, G. Explaining and forecasting online auction prices and their dynamics using functional data analysis. *Journal of Business and Economic Statistics*, in press, 2006.
131. Ward, S. G., and Clark, J. M. Bidding behavior in on-line auctions: An examination of the eBay Pokemon card market. *International Journal of Electronic Commerce*, 6, 4, 2002, 139-155.
132. Wilcox, R. T. Experts and amateurs: The role of experience in Internet auctions. *Marketing Letters*, 11, 2000, 363-374.
133. Wood, C. A., Fan, M., and Tan, Y. An examination of the reputation systems for online auctions. Working paper, Mendoza College of Business, University of Notre Dame, Notre Dame, IN, 2003.
134. Wood, C. A., and Ow, T. WebView: A SQL extension for joining corporate data to data derived from the Web. *Communications of the ACM*, 48, 9, 2005, 99-104.
135. Yamagishi, T. The role of reputation in open and closed societies: An experimental study of Internet auctioning. Working paper, Department of Behavioral Science, Graduate School of Letters, Hokkaido University, Sapporo, Japan, 2003.
136. Yin, P. Information dispersion and auction prices. Working paper, Harvard Business School, Harvard University, Cambridge, MA, 2002.

APPENDIX. AGENT-BASED DATA COLLECTION

Some natural questions will be asked related to this research that are worthwhile to briefly consider. They include: When a researcher is interested in developing software agents for Internet-based data collection, what are the key design issues? How do the design issues relate to the likely success a researcher will have with the use of the agent-based data collection methodology? What advice is available in the current research literature to guide this kind of work, and where might a novice begin? In this Appendix, we will briefly answer these questions, and point the reader to the appropriate literature.

A good starting point for the interested reader is an article by Kauffman et al. [67] on the design of *long-lived Internet agents*. The primary contrast with respect to data-collecting software agents lies in the length of time that they operate on behalf of the user. Consumers and buyers typically employ transaction agents to retrieve small amounts of data related to transaction events. Managers and researchers are more likely to use long-lived Internet agents, which provide the capability to retrieve, store, process and analyze large amounts of data over a long period of time. When undertaking a project involving Internet-based agent-led data collection, it is important for the researcher to recognize the issues and difficulties that are associated with this approach. Kauffman et al. [67] noted five issues that must be appreciated. First, there still is no single standard for the presentation of data on Web pages on the Internet. Second, HTML has no standard elements related to the underlying data, and so it does not readily map to what is needed for effective data collection agent design. XML is somewhat better in that data definitions can be discovered or created, however, in practice there has not been enough effort put towards the standardization of XML data in different business and e-commerce contexts. Third, conducting longitudinal data collection requires some tolerance to the typical faults that occur in the natural conduct of organizational and business processes involving computers, including Web site downtime, network and system crashes, data with errors, high usage loads and so on – all of which result in faults, outages and an impaired ability to conduct high quality automated data collection.

Fourth, part of the problem that the researcher will face involves identifying the Web pages from which to extract data. This consideration implies that appropriate software for the purpose of Internet-based data collection should also have the capacity to conduct search and navigation. Fifth, even though many, many Web servers are available on the Internet now – so the places from which one could possibly collect data are endless – nevertheless, researchers must have some sensitivity to the propriety and impact of their automated search-and-extract activities on the Internet. Allen et al. [1] point out that there is yet to be established a code of conduct for data collection on the Web – what's admissible, what's not, and how to determine what available data and information should properly be treated as private property.

These issues are likely to affect the success that researchers will have in extracting the kind of re-

search data that they wish to obtain from the Internet. Kauffman et al. [67, p. 219] note that recovery from an abnormal terminal for a data collection agent on the Internet typically requires context sensitivity and varies according to the history of work previously performed. So data collection agents need to demonstrate *persistence*, in that they are not only able to restart their processes of data collection, but also do so in a manner that restores the state of the software agent prior to the abend. Kauffman et al. [67, p. 220] further point out that Internet agents are *data-centric* in that [t]hey must acquire, interpret and transfer data, possibly modifying their behavior based on the data that they acquire.

These observations led the authors to argue that there are five dimensions of *agent sophistication* necessary (beyond the typical performance requirements for good computer programs) that Internet agents must demonstrate. They include intelligence, concurrency, validation, recovery, monitoring, and interactivity. *Intelligence* is the artificial intelligence-like ability to use rules to mimic the responses of humans, and for the software agent to identify the conditions associated with its data extraction query, and adjust its data collection targets and behavior accordingly. *Concurrency* in a software agent is the capability to do several things at once, and is implemented through multi-threaded processes that make it possible to take advantage of available CPU time. *Validation* capability in an Internet agent is the means to ensure that the data are of the proper type and identity, and match the rules associated with their extraction. *Recovery* is the capacity of a software agent to recover from problems that cause data collection to fail. Klein and Dellarocas [75] have pointed out that it is burdensome to program agents with exception handling code, and that it may help to rely on generic recovery routines that can be called upon as reusable external software objects. *Monitoring* is the capability of a software agent to periodically track events that give rise to new data on Web sites that are targeted for data collection. The capability is in contrast to the one-and-done approach; scheduled monitoring and continuous monitoring are both possible. Finally, *interactivity* is a software agent's ability to modify its data collection behavior based on the data it collects.

Kauffman et al. [67] provide an interesting demonstration of a tool for Web-based data collection called, eDRILL, an *electronic data retrieval lexical agent*. Their demonstrational application is to eBay electronic auctions for collectible coins. The authors placed the agent sophistication constructs in the context of the Unified Modeling Language (UML), as a means to effectively express the key design concepts that go along with the dimensions of agent sophisticated. Overall, however, these capabilities suggest that truly longitudinal data collection-based research designs on the Internet will be subject to multiple threats. They include missing data, incompletely collected data, data collected under inappropriately specified conditions, invalid data, and other problems that lie in the domains of the capabilities of the software agent and its interaction with a dynamic computing and networking environment. Other issues that

may arise involve changes in Web site designs made by Internet-based sellers, requests to cease and desist from data collection at a particular site, difficulties in manipulating the very large gigabyte databases that are collected to extract the appropriate data (a special problem with logfiles), and so on.

We would like to recommend three additional resources that may be helpful for beginners in their use of this means of data collection. The most basic things to know about in this area are some software tools that will help to extract material from Web pages. An example is ListGrabber (www.egrabber.com), and other associated software, which permits a user to identify the kinds of data that are targeted for collection as lists. Although the software is sold with an emphasis on grabbing contact information, resumes and response data from the Web, it can be adapted to pull list data and port it to Excel, so it then can be further massaged and cleaned to port to database software and to statistics packages. This tool works directly with HTML. A second example involves the extraction of data from Extensible Markup Language (XML) pages on the Web. Microsoft Access 2002 and 2003 have had the capability to support importing XML. Microsoft states that [w]ithout XML, this task might involve exporting the data [from a Web site] to a text-based file (assuming that the various data sources supported this), manipulating the data files by adding delimiters to separate the data into discrete parts, importing the data into SQL Server or Access, and then spending a considerable amount of time cleaning up the data. Using XML allows you to minimize this kind of time spent reformatting and cleaning up your data (office.microsoft.com/en-us/access/HA010345601033.aspx?pid=CL100570041033).

Wood and Ow [134] offer the another useful source for beginners to gain specific insights into the technical side of the process of Web-based data collection based on an article published in the *Communications of the ACM*. They focus on extracting Web-based data using SQL extensions and a tool called WebView that supports open database connectivity (ODBC) for database programs such as Microsoft Access, Oracle, and SQL Server. Wood and Ow's innovation is to create the basis for dynamic Web views for Web page data in a relational database joined with dynamic organizational and user views of a corporate database. This innovation is founded on prior work by Lakshmanan et al. [76]. These authors identified a means to achieve database interoperability through the creation of SchemaSQL. This is an extension of SQL that permits queries to database schemas and databases, the transformation of data in database structures that are different from the original database, the creation of database views that are dependent on input values related to the query, and additional facilitation of relational database interoperability and data/schema manipulation. Wood and Ow [134, pp. 101-102] note five requirements for an effective WebView SQL extension: (1) it must have expressive power that is independent of HTML, XML, or other Web-based markup languages; (2) it must allow the restructuring of Web data to conform to a database schema; (3) it must be shown to be sufficient to capture any Web data, including XML or

HTML; (4) it must function like existing database constructs to allow transparency for the database developer; and (5) it must be efficiently implemented. The interested reader should see this article for additional details on an application of these ideas to data collection in eBay auctions.

BRIEF BIOS OF THE AUTHORS

ROBERT J. KAUFFMAN is currently the W.P. Carey Chair in Information Systems at the W.P. Carey School of Business, Arizona State University. He previously was Director of the MIS Research Center, and Professor and Chair in the Information and Decision Sciences Department at the Carlson School of Management, University of Minnesota (1994-2007). Rob previously worked in international banking, and served on the faculty of New York University (1988-1994) and the University of Rochester (1992), before moving to Minnesota in 1994. His graduate degrees are from Cornell University (M.A., 1979) and Carnegie Mellon University (Ph.D., 1988). His current research focuses on senior management issues in IS strategy and business value, financial evaluation and technology investments, technology adoption, e-commerce and electronic markets, pricing strategy and supply chain management issues. His research has been published in *Organization Science*, *Journal of Management Information Systems*, *Communications of the ACM*, *Management Science*, *MIS Quarterly*, *Information Systems Research*, *Decision Sciences*, and other leading IS, economics and computer science journals and conferences. He recently won outstanding research awards with his doctoral students and faculty colleagues at the INFORMS Conference on IS and Technology in 2003, 2004 and 2005, the Hawaii International Conference on Systems Science in 2004, and the International Conference on Electronic Commerce in 2005. In 2006, he won an outstanding research contribution award from the IEEE Society for Engineering Management for an article on standards drift in technology adoption in the *IEEE Transactions on Engineering Management*.

Charles A. Wood is currently an Assistant Professor of Management at the Mendoza School of Business, Notre Dame University. He is a graduate of the University of Minnesota's Carlson School of Management, where he received a doctorate in Information Systems in 2001. His research interests lie in the area of electronic commerce, electronic auctions, buyer and seller reputation, and pricing in Internet-based selling. He received a best paper award in 2000 for his research on follow-the-leader pricing dynamics in e-commerce. His research publications have or will appear in *Management Science*, *Information Technology and Management*, *Electronic Commerce Research and Applications*, *Managerial Decisions and Economics*, and the *International Journal of Intelligent Systems in Accounting, Finance and Management*. He has presented his research at leading conferences in the fields of IS, Marketing, Computer Science and Economics, and is the author of multiple books on various topics in software development, database management and computer programming.