ELSEVIER
JAI

RESEARCH METHODOLOGY IN STRATEGY AND MANAGEMENT

VOLUME 2

DAVID J. KETCHEN, JR.
DONALD D. BERGH
Editors

RESEARCH METHODOLOGY IN STRATEGY AND MANAGEMENT

RESEARCH METHODOLOGY IN STRATEGY AND MANAGEMENT VOLUME 2

RESEARCH METHODOLOGY IN STRATEGY AND MANAGEMENT

EDITED BY

DAVID J. KETCHEN, JR.
College of Business, Florida State University, USA

DONALD D. BERGH
Krannert Graduate School of Management, Purdue University, USA

2005

ELSEVIER
JAI

Amsterdam – Boston – Heidelberg – London – New York – Oxford
Paris – San Diego – San Francisco – Singapore – Sydney – Tokyo

ELSEVIER B.V.
Radarweg 29
P.O. Box 211
1000 AE Amsterdam
The Netherlands

ELSEVIER Inc.
525 B Street, Suite 1900
San Diego
CA 92101-4495
USA

ELSEVIER Ltd
The Boulevard, Langford
Lane, Kidlington
Oxford OX5 1GB
UK

ELSEVIER Ltd
84 Theobalds Road
London
WC1X 8RR
UK

First edition 2005

British Library Cataloguing in Publication Data
A catalogue record is available from the British Library.

ISBN: 0-7623-1208-4
ISSN: 1479-8387 (Series)

∞ The paper used in this publication meets the requirements of ANSI/NISO Z39.48-1992 (Permanence of Paper).
Printed in The Netherlands.

CONTENTS

LIST OF CONTRIBUTORS

Jay B. Barney	Fisher College of Business, The Ohio State University, Columbus, USA
Donald D. Bergh	Krannert School of Management, Purdue University, West Lafayette, USA
Philip Bromiley	Carlson School of Management, University of Minnesota, Minneapolis, USA
James G. Combs	College of Business, Florida State University, Tallahassee, USA
Joseph E. Coombs	Robins School of Business, University of Richmond, USA
T. Russell Crook	College of Business Administration, Northern Arizona University, Flagstaff, USA
Catherine M. Dalton	Kelley School of Business, Indiana University, Bloomington, USA
Dan R. Dalton	Kelley School of Business, Indiana University, Bloomington, USA
Timothy B. Folta	Krannert School of Management, Purdue University, West Lafayette, USA
Devi R. Gnyawali	Pamplin College of Business, Virginia Polytechnic Institute and State University (Virginia Tech), Blacksburg, USA
R. Duane Ireland	Mays Business School, Texas A&M University, College Station, USA
Scott Johnson	William S. Spears School of Business, Oklahoma State University, Stillwater, USA
David J. Ketchen, Jr.	College of Business, Florida State University, Tallahassee, USA
Michael H. Lubatkin	School of Business, University of Connecticut, Storrs, USA

Tyson B. Mackey	Fisher College of Business, The Ohio State University, Columbus, USA
Kent D. Miller	The Eli Broad Graduate School of Management, Michigan State University, East Lansing, USA
Margaret Peteraf	Tuck School of Business at Dartmouth, Hanover, USA
Mark Shanley	Krannert School of Management, Purdue University, West Lafayette, USA
Christopher L. Shook	Auburn University, Auburn, USA
Zeki Simsek	School of Business, University of Connecticut, Storrs, USA
Caron H. St. John	Clemson University, Clemson, USA
Beverly B. Tyler	College of Management, North Carolina State University, Raleigh, USA
John F. Veiga	School of Business, University of Connecticut, Storrs, USA
Justin W. Webb	Mays Business School, Texas A&M University, College Station, USA

INTRODUCTION

Welcome to the second volume of Research Methodology in Strategy and Management. This book series' mission is to provide a forum for critique, commentary, and discussion about key research methodology issues in the strategic management field. Strategic management relies on an array of complex methods drawn from various allied disciplines to examine how managers attempt to lead their firms toward success. The field is undergoing a rapid transformation in methodological rigor, and researchers face many new challenges about how to conduct their research and in understanding the implications that are associated with their research choices. For example, as the field progresses, what new methodologies might be best suited for testing the developments in thinking and theorizing? Many long-standing issues remain unresolved as well. What methodological challenges persist as we consider those matters? This book series seeks to bridge the gap between what researchers know and what they need to know about methodology. We seek to provide wisdom, insight, and guidance from some of the best methodologists inside and outside the strategic management field.

Before we discuss the contents of this volume, let us briefly reflect on its predecessor. Volume 1 debuted at the 2004 Academy of Management meeting in New Orleans, Louisiana. The volume was showcased in a symposium sponsored by the Business Policy and Strategy division at the Academy. We were surprised and delighted that our large meeting room was filled to capacity. We believe this turnout reflects the desire of strategy researchers to improve their methodology skills; a desire we hope this book series will serve capably. We want to thank Hüseyin Tanriverdi, Margarethe Wiersema, Pam Barr, Kevin Carlson, Don Hatfield, and Larry Williams for offering excellent presentations of their chapters in New Orleans. In addition to the symposium, three copies of the book were available at the Elsevier promotional booth for examination by our peers. On a somewhat bizarre note, one of the display copies was stolen from the booth. The display copy at the 2004 Strategic Management Society meeting was stolen as well. We are not sure if we should be flattered or appalled. Perhaps both!

We hope that Volume 2 also will be well received. The volume you hold in your hands offers 11 diverse chapters that can be categorized roughly into three sets. One set of chapters describes challenges and opportunities

inherent in particular content areas. The opening chapter by Jay B. Barney and Tyson Mackey offers insightful guidance about testing resource-based theory. This theory has become perhaps the dominant perspective in strategic management research following the publication of Barney's seminal 1991 *Journal of Management* article. We are thrilled to provide a forum for Jay and his co-author to address the methodology side of the theory; an aspect that has received far less attention in the literature than the conceptual side. Mark Shanley and Margaret Peteraf tackle the daunting (and perhaps unenviable) task of trying to move methodology forward within strategic groups research. These authors are well armed for this task, however, as their articles on strategic groups and competition in general are among the most highly regarded in the literature.

The next chapter focused on a content area takes seriously the "and Management" portion of our book series' title. Duane Ireland, Justin Webb, and Joe Coombs enlighten us about theory and methodology in entrepreneurship research. There are parallels between the state of the entrepreneurship area today and the state of strategic management research a couple of decades ago. Just as with strategy in the 1980s, entrepreneurship tackles issues of practical importance, but critics question its theoretical robustness and methodological rigors. Led by one of the premier scholars bridging strategy and entrepreneurship, this chapter offers practical and thought-provoking advice.

Real options is a concept that has generated considerable discussion among strategy researchers in recent years. One of the leading scholars on the topic, Tim Folta, poses and addresses the provocative question of "are real options real?" Finally, Jim Combs, Russell Crook, and Chris Shook examine the dimensionality of the concept of organizational performance. They evaluate the construct validity of performance and then report meta-analyses that show implications of variations in approaches used to measure performance. Although performance is the key dependent variable in many strategy research streams, there has not been a thorough deconstruction of the concept since Venkatraman and Ramanujam's landmark *Academy of Management Review* article in 1986. Thus, Combs et al.'s focus is timely and warranted.

A second set of chapters adopts a reflective, philosophical slant. They examine key ontological and epistemological issues in the strategic management context. First, Bromiley and Johnson discuss the relationship between method and theory. They argue that researchers may not be testing theories when they fail to specify the underlying mechanisms of the theory of interest. Examples from the upper echelons, diversification, and transaction-cost

economics literatures are used to demonstrate how researchers can fail to actually test theories accurately. Bromiley and Johnson provide insights into how the integration between method and theory can be improved.

Next, Kent Miller discusses the personal and human side of research methods. He draws from philosophical approaches to consider how human capacities influence how we, as people, conduct research. Miller identifies and discusses how researcher orientation influences personal judgments, methodological pluralism, and social practices. We believe you will find his insights intriguing. Caron St. John provides a comprehensive overview of key issues in mixed-level strategy research. She explains how strategic management researchers tend to conduct multi-level research and identifies common problems that arise and what implications those tend to have for theory and knowledge development. She discusses the conventional practices relative to dominant theories and views in strategic management research and provides directions for improving how researchers approach cross- and multi-level research. Overall, the chapter offers a great primer for those interested in spanning levels of analysis in their inquiry.

A final set of chapters considers the use of specific methodological techniques and offers specific suggestions that researchers can implement to strengthen their studies. Dan Dalton and Catherine Dalton have been prominent users of meta-analysis in recent years. The technique holds great promise for strategic management, but it is used less frequently in our field than in allied disciplines such as organizational behavior. Dalton and Dalton highlight the key action issues for using meta-analysis to aggregate strategy studies. Their straightforward, practical advice should facilitate greater use of this powerful technique.

The emergence and growth of the Internet has produced new opportunities for researchers. Zeki Simsek, John Veiga, and Michael Lubatkin examine one of the most fertile: using the Internet to conduct surveys. Internet-based surveys can cut costs dramatically and can reach thousands of potential respondents in seconds. Yet, unique challenges such as how to ensure one is gathering a representative sample are created. Simsek et al. offer a series of suggestions to guide the Digital Age survey researcher in navigating such issues. Finally, Devi Gnyawali and Beverly Tyler examine cognitive mapping techniques, with a particular emphasis on cause mapping. Their methodologies focus on strategic judgment processes and offer new insights into the 'black box' of strategic decision-making. The authors provide a comprehensive overview of cause mapping models and identify the issues that arise within this perspective. Importantly, they provide guidance on how to improve the application of cause-mapping techniques.

Overall, the 11 chapters contained in this volume attempt to gently nudge the field toward better practice. If their recommendations are followed, the result will be an enhanced ability to understand how organizations act and perform. We hope your research benefits from these chapters as much as we enjoyed working with their respective authors. We are very grateful to all of the contributors for their insights and efforts.

David J. Ketchen, Jr.
Donald D. Bergh
Editors

TESTING RESOURCE-BASED THEORY

Jay B. Barney and Tyson B. Mackey

ABSTRACT

While strategy scholars once thought that the resource-based view could not be tested directly by observing resources, recent work has dispelled this notion. While resources are difficult to measure, many clever scholars have been able to measure resource heterogeneity and performance.

In 1916, Albert Einstein predicted the existence of gravity waves. Given how small these waves were supposed to be – 10^{-18} of a millimeter – Einstein was convinced that this implication of his general theory of relativity would never be examined directly. Initially, the existence of these waves was only examined indirectly, by observing that pulsars were losing mass at a rate consistent with the existence of gravity waves. However, more recently, a new generation of wave detection technology has been introduced. Drawing on the computing power of thousands of personal computers linked in a voluntary network, physicists now believe that it may be possible to directly observe gravity waves, although it may take many years to refine the technology and complete the data analysis (Lafferty, 2005).

Godfrey and Hill (1995) observed that resource-based theory – along with transaction cost economics and agency theory – incorporated difficult to

Research Methodology in Strategy and Management
Research Methodology in Strategy and Management, Volume 2, 1–13

ISSN: 1479-8387/doi:10.1016/S1479-8387(05)02001-1

observe concepts as independent variables. These authors wondered if it would ever be possible to directly test resource-based theory. Initially, they reasoned, resource-based empirical work would have to focus on examining the observable implications of a firm's resources and capabilities, rather than examining those resources directly. However, more recently, several scholars have begun to develop techniques for measuring at least some aspects of these previously difficult to observe concepts. Although it may take many years to refine this measurement technology and complete the data analysis, there is now a growing belief that it may be possible to measure resources and capabilities and therefore to directly test the implications of resource-based theory (Dutta, Narasimhan, & Rajiv, 2005).

In reviewing these stories, in no way is it being suggested that resource-based theory has the same theoretical status as Einstein's theory of general relativity. Rather, these stories are reviewed only to point out that the evolution of science – whether it is experimental physics or empirical social science – often involves the development of new approaches to measurement and testing that make what were once impossible to test theories testable. Indeed, in the ever-growing literature that now constitutes the "resource-based view," a great deal has been learned over the last several years about how to test this theory (Barney & Arikan, 2001). The purpose of this chapter is to highlight some of these lessons.

THE QUESTION OF VALUE

It is now widely understood that resources – the tangible and intangible assets controlled by a firm that enable it to create and implement strategies (Barney, 2002) – only have the potential to generate economic value if they are used to do something (Porter, 1991). Of course, the thing that resources – and their close conceptual cousin, capabilities (Amit & Schoemaker, 1993) – are supposed to do is to enable firms to create and implement strategies.

This simple insight actually suggests a way that researchers can measure the potential of a firm's resources to create value: To measure this potential, measure the value created by the strategies a firm creates and implements using its resources. Put differently, since resources have no value in and of themselves and only create value when they are used to implement strategies, researchers should examine the value these strategies create to infer the potential value of a firm's resources.

Of course, there is substantial literature that describes the ability of different strategies to create economic value. A wide variety of such strategies

has been described, including cost leadership, product differentiation, vertical integration, flexibility, tacit collusion, strategic alliances, corporate diversification, mergers and acquisitions, and international strategies, to name just a few (Barney, 2002). Much of this work identifies the conditions under which these strategies will and will not create economic value.

For example, a cost leadership strategy creates value if and only if it enables a firm to reduce its costs below those of competing firms (Porter, 1980). A product differentiation strategy creates value if and only if it enables a firm to charge higher prices for its products than a firm that is not differentiating its products (Porter, 1980). A corporate diversification strategy creates value if and only if it exploits an economy of scope that cannot be realized through market contracting (Teece, 1980).

There has been less work that links specific firm resources and capabilities with the ability to create and implement these kinds of firm strategies. This is largely because currently available typologies of firm resources are very broad in scope, e.g., Barney's (2002) distinction between financial, physical, human, and organizational resources. Further work developing this type of typology is likely to facilitate the examination of the link between resources, in general, and the ability to conceive of and implement specific strategies.

However, that there has been limited work that links specific resources to particular strategies does not mean that there has been no work in this area. Indeed, several papers have examined the linkages between particular resources and capabilities and specific strategies. Most of this work is carried out on a limited sample of firms within a single industry. This helps establish the link between the resources and strategies in question. But, taken as a whole, this work suggests an approach to linking resources to strategy and thereby examining the potential of resources to create economic value by enabling firms to create and implement strategies. Consider a couple of examples of this research.

In 1994, Henderson and Cockburn were interested in understanding why some pharmaceutical firms were more effective in developing new patentable drugs than other pharmaceutical firms. It is well known that patents are a source of economic value in the pharmaceutical industry (Mansfield, Schwartz, & Wagner, 1981) – contingent on the demand for particular drugs, firms with large numbers of patented drugs will usually have higher revenues than firms with smaller numbers of patented drugs. The specific resource that Henderson and Cockburn were able to identify that enabled some firms to have more patents than other firms was something they called "architectural competence" – the ability to facilitate cooperation among the different scientific disciplines required to develop and test a new

pharmaceutical drug. Firms with high levels of this competence were able to patent more drugs than firms with low levels of this competence. Henderson and Cockburn's research showed that architectural competence had the potential to generate economic value when it was used to develop new patentable drugs.

More recently, Ray, Barney, and Muhanna (2004) examined the relationship between the ability of two functional areas – the information technology function and the customer-service function – and the level of customer service in a sample of North American insurance companies. Again, it is widely recognized that customer service is an information intensive function in most modern insurance companies, and that the careful use of information technology can enhance the ability of customer-service professionals to meet their customers' needs. Customer satisfaction, in turn, is related to a variety of economically important variables, including customer retention. What Ray et al. (2004) were able to do is to develop a measure of the level of cooperation between the IT and customer-service functions in a sample of insurance firms and demonstrate that this relationship – a socially complex resource – has the potential to create economic value when it is used to develop customer-service-specific IT applications.

Besides demonstrating that it is possible to examine the potential of a resource to create economic value by examining the value consequences of the strategies a firm creates and implements by using these resources, these – and related – papers have several other things in common. First, they are examples of what might be called "quantitative case studies." That is, they examine the relationship between a firm's resources and the value of its strategies in a narrow sample of firms, typically a sample of firms drawn from a single industry. This enabled these authors to clearly identify industry-specific resources and capabilities and to build industry-specific measures of these resources. Then, using traditional quantitative techniques, they examine the relationship between these measures of firm resources and attributes of a firm correlated with a firm's economic performance.

Of course, it is difficult to generalize this research beyond the specific industry contexts within which it is done. That architectural competence is related to the number of patents in pharmaceutical firms may or may not say anything about the relationship between architectural competence and innovation in other firms in other industries. That the level of cooperation between IT and customer service has an impact on the level of customer service in North American insurance companies may or may not say anything about the relationship between this type of cooperation and customer service in other firms in other industries.

Although these papers have limited generality at the level of the specific resources and strategies studied, their results are quite general from a broader perspective. Each of these papers – and the several others that apply a similar empirical logic (e.g., Combs & Ketchen, 1999) – show that at least some firm resources have the potential to generate economic value if they are used to create and implement certain strategies. Over time, as more of these quantitative case studies are done, our ability to specify the conditions under which resources can be used to create and implement strategies that create economic value will be enhanced.

Second, many of these studies examine the value potential of a firm's resources at a level of analysis below that of the firm. Not surprisingly, the most correct level of analysis at which to examine the relationship between a firm's resources and its strategies is at the level of the resource, not the level of the firm. However, the firm is usually the unit of accrual. We are likely to learn a great deal more about the relationship between resources and strategies if scholars are able to "get inside" the firm, where resources reside, rather than simply correlate aggregate measures of resources with aggregate measures of the value of a firm's strategies (Rouse & Daellenbach, 1999).

This, of course, implies that the best resource-based empirical work will involve collecting primary data from within firms in a carefully drawn sample. The norm in much of the currently published work in strategic management seems to be to use publically available data sets to test the extant theory. Clearly, some very clever scholars have been able to use these data sets to say some interesting things, even about resource-based theory. One example is the Miller and Shamsie (1996) study of resources and performance in the motion picture industry. The creative use of proxies helped this article win the *Academy of Management Journal's* best paper award for 1996. However, in the long run, going inside a sample of firms and collecting data about resources and strategies directly seems likely to be more important for the development and evolution of resource-based research. Even seemingly unobservable resources may be assessed by going inside firms. For example, Hult and Ketchen (2001) used measures gathered from informants and the latent construct function of structural equation modeling to tap into intangible resources.

Finally, the central independent variables in both of these papers – architectural competence in Henderson and Cockburn (1994) and IT/customer-service cooperation in Ray et al. (2004) – focus on a particular type of organizational resource. This type of resource has been described as socially complex (Barney, 1991) and it has been linked to the sustainability of a

firm's competitive advantage. Empirically, examination of these sustainability issues is carried out in the next section of this chapter.

SUSTAINING COMPETITIVE ADVANTAGES

It is now widely understood that resources only have the *potential* to create economic value, and that the potential is only realized when a firm uses its resources to create and implement strategies. It is perhaps not as widely recognized that the ability of other firms to imitate a particular firm's strategies does not depend on the attributes of those strategies, per se, but rather on the attributes of the resources and capabilities that enabled that firm to create and implement its strategies in the first place. Put differently, just as resources only have the potential to create value through their impact on a firm's strategies, so too strategies only have the potential to be costly to imitate because of the nature of the resources that enabled a firm to choose and implement its strategies.

By their nature, strategies are relatively public. That is, when a firm implements its strategies, it is usually not very long before other firms are able to articulate what those strategies are. This is especially the case when a firm's strategies are logical and coherent.[1] What are not always so public are the resources and capabilities that enable a firm to create and implement its strategies.

Resource-based theory suggests that valuable strategies that are created and implemented using resources that are widely held or easy to imitate cannot be a source of sustained competitive advantage (Barney, 1991). In this context, a firm has a sustained competitive advantage when it is one of only a few competing firms that is implementing a particular value creating strategy and when this competitive situation lasts over extended periods of time.

Resource-based theory also makes specific predictions about the characteristics of resources and capabilities that make some more difficult to imitate than others.[2] For example, Lippman and Rumelt (1982) suggest that causally ambiguous resources are more likely to be costly to imitate than resources that are not causally ambiguous. Barney (1986a) suggests that resources and capabilities a firm already controls are more likely to be costly to imitate than resources it acquires from competitive factor markets. Barney (1986b) suggests that socially complex resources and capabilities – the particular resource he examined in this paper was a firm's culture – are more likely to be costly to imitate than resources that are not socially

complex. Dierickx and Cool (1989) suggest that resources characterized by time compression diseconomies, asset stock interconnectedness, and asset mass efficiencies are more likely to be costly to imitate than resources without these attributes. Finally, in a summary, Barney (1991) suggests that path-dependent, causally ambiguous, and socially complex resources are more likely to be costly to imitate than resources without these attributes.

Of course, each of these assertions implies testable hypotheses about the imitability of different types of resources. A study that examined, say, path-dependent resources that enabled a few competing firms to create and implement value-creating strategies, but where numerous firms were able to imitate these strategies once they were initially implemented, would be very inconsistent with resource-based theory. So too would a study that examined resources that did not possess any of these special attributes but nevertheless enabled a few competing firms to create and implement value creating strategies, but where numerous firms were *unable* to imitate these strategies once they were initially implemented. In the first study, path dependence would not be a source of sustained competitive advantage; in the latter study, the lack of path dependence (or social complexity, or causal ambiguity, or some other attribute of resources supposed to prevent their easy imitation) would be a source of sustained competitive advantage. Both results contradict resource-based theory.

Of course, the empirical requirements to test these hypotheses are non-trivial. But several studies have come close to approximating these requirements. For example, by studying the resource-based determinants of patents, Henderson and Cockburn (1994) come close to examining the sustainability of any competitive advantages that architectural competence might create because patents, as a function of patent law, last a defined and relatively long period of time – 20 years.[3]

One particularly elegant study that examined the imitability of path-dependent firm resources was published by Barnett, Greve, and Park (1994). In this paper, Barnett et al. examined why some commercial banks competing in the state of Illinois during a recession were able to outcompete other banks competing in the same market at the same time. Clearly, banks that were not performing well in this setting had a very strong incentive to imitate the strategies of banks that were performing well. However, Barnett et al. hypothesized that one reason the strategies of the banks that were doing well were not subject to quick imitation was that these banks possessed resources and capabilities that enabled them to choose these valuable strategies, and that these underlying resources and capabilities were costly to imitate due to their path-dependent nature.

This study did not directly measure the resources that enabled some banks to outperform other banks. However, it did demonstrate that banks that had survived a financial recession previously systematically outperformed banks that had not survived a financial recession previously. Barnett et al. interpreted this finding to suggest that there was something about the historical experience of banks that had survived a previous recession that had equipped them with the resources and capabilities – they use the largely interchangeable term "routines" (Nelson & Winter, 1982) – necessary to survive, and even prosper, in a later recession. Of course, this paper would have been even stronger if it could have directly measured these resources and the extent to which they were path-dependent in nature. Nevertheless, it is consistent with the general hypothesis that path-dependent resources and capabilities are costly to imitate and thus a source of sustained competitive advantage.

Makadok's (1999) study of economies of scale in the money-market mutual fund industry also supports resource-based assertions. In this case, however, resource-based theory would suggest that since the realization of these economies of scale did not depend on resources or capabilities that are costly to imitate, the strategies that exploit these economies of scale would not be a source of sustained competitive advantage for these firms. If Makadok (1999) had found that economies of scale in this industry had been a source of sustained competitive advantage, this would have been inconsistent with resource-based theory.

Interestingly, these two studies, like the first two studies reviewed in this chapter, are quantitative case studies. That is, they studied a sample of firms drawn from a particular industry, and in the case of Barnett et al. (1994), from a particular geographic market. This enabled these scholars to examine the link between specific resources, strategies, and competitive advantage over time.

Unfortunately, neither of these studies measured the attributes of a firm's resources and capabilities directly. This is, perhaps, due to the difficulty of gaining access to this intra-organizational resource-level information over an extended period of time. Obviously, duplicating Ray et al. (2004) survey methodology over several years would be very challenging and would delay the publication of any subsequent paper until after all the data had been collected. A recent paper by Leiblein and Miller (2003) on transaction cost and resource-based implications for vertical integration decisions comes closer to meeting this ideal standard than much of the previous work on sustainability of competitive advantages.

Another attribute shared by these two studies is that they were conducted on data over time. Although it is possible to define sustained competitive advantage with respect to the observed inability of firms to imitate a particular firm's resources (Barney, 1991), this equilibrium definition of sustained competitive advantage will often be highly correlated with competitive advantages that last for a long time. This suggests that time-series analyses of various kinds will generally be required to investigate the imitability of different types of firm resources and thus the sustainability of a firm's competitive advantage. The challenges associated with collecting resource-level information within a firm over time have already been discussed.

THE QUESTION OF ORGANIZATION

Thus, though much work is left to be done, some research has examined what most consider to be the two central assertions of resource-based theory: (1) that some resources have the potential to enable firms to create and implement valuable strategies and (2) that such resources can be a source of sustained competitive advantage when they possess attributes that make their imitation costly. However, some versions of resource-based theory also suggest that firms must be organized to take advantage of their resources and strategies if their full economic potential is to be realized (Barney, 2002). This emphasis on strategy implementation has received less attention in the resource-based empirical literature.

There are several possible reasons for this relative inattention. First, most strategy scholars are interested in understanding sources of sustained competitive advantage. If a firm's ability to implement strategies is valuable (in the sense described earlier), rare, and costly to imitate, then a firm's strategy implementation capability is a potential source of sustained competitive advantage. In this case, the study of strategy implementation – as a source of sustained competitive advantage – is indistinguishable from other studies of the sources of sustained competitive advantage.

Indeed, some of the studies reviewed thus far could easily be reinterpreted as if they were examining the competitive consequences of a firm's ability to implement its strategies. Thus, the Henderson and Cockburn (1994) study could be reinterpreted as a strategy implementation study by suggesting that architectural competence is the ability that some firms have to implement their patenting strategies more effectively than other firms. In this sense, because the ability to implement a strategy can be thought of simply as

another type of resource or capability, strategy implementation can be thought of as just another possible source of sustained competitive advantage.

This is one reason why research on the ability of firms to develop new capabilities – so called "dynamic capabilities" (Teece, Pisano, & Shuen, 1997) – has captured the interest of so many strategy scholars. Such dynamic capabilities can also be reinterpreted in strategy implementation terms: A dynamic capability is the ability that some firms have to create new capabilities, capabilities whose potential value can only be realized when a firm implements new strategies that build on these new capabilities.

However, another perspective on the question of organization is that organization includes all those dimensions of implementing a firm's strategies that are, in principle, imitable, but are nevertheless important if a firm is to gain competitive advantages. Barney (2002) calls these dimensions of strategy implementation "complementary resources," because these implementation skills – things like an organization's structure, its management controls, and its compensation policies – are not sources of competitive advantage by themselves, but are nevertheless important if a firm realizes the full competitive potential of its resources and strategies.

Some research has focused on the imitability of these strategy implementation skills. For example, Armour and Teece (1978) examined the impact of the M-form organizational structure on firm performance. Resource-based theory suggests that such a structure, by itself, should not be a source of sustained competitive advantage. And indeed, Armour and Teece show that the M-form structure was a source of advantage for the first few firms that implemented it but was not a source of competitive advantage over time.

Although a few examples of this kind of work can be cited, it must be admitted that relatively little work has examined the competitive implications of strategy implementation skills that are in principle imitable. This is probably because, as Armour and Teece showed, these kinds of capabilities are not likely to be sources of sustained competitive advantage. However, the fact that these capabilities are, in principle, imitable does not necessarily mean that they will be widely imitated. Indeed, an interesting unanswered question facing resource-based theorists is: Why are highly imitable resources that enable firms to choose and implement economically valuable strategies not imitated, even by firms that stand in a competitive disadvantage to firms with these resources? This and related questions have not received the attention they deserve in the empirical resource-based literature.

CONCLUSION

The number of empirical tests of resource-based theory continues to grow rapidly. For example, a recent issue of the *Strategic Management Journal* (SMJ) (Vol. *26*(3), March 2005) included three articles that examined implications of resource-based theory. Two of these articles (Song, Droge, Hanvanich, & Calantone, 2005; Dutta et al., 2005) apply quite sophisticated estimation techniques. Song et al. (2005) apply their techniques to data collected from a survey taken from a sample of joint ventures; Dutta et al. (2005) apply their techniques to the analysis of R&D capabilities of firms within the U.S. semiconductor industry. In this sense, these two articles are examples of the quantitative case studies described earlier in this chapter.

What is most striking about this recent issue of *SMJ* is not that it is unusual, but rather, that it is typical of this journal over the last several years. Virtually every issue of *SMJ* has included one, or several, empirical tests of resource-based theory. Some apply relatively simple estimation techniques (e.g., Ray et al., 2004), others apply quite sophisticated approaches (e.g., Song et al., 2005, Dutta et al., 2005; Hansen, Perry, & Reese, 2004; Hatch & Dyer, 2004). However, most adopt some version of the methodological approaches described here, including quantitative case studies, data collected at the resource unit of analysis, data collected over time, and so forth.

Scholars continue to ask, "How does one measure resources?" Usually, the question they are really asking is "How does one measure resources, easily?" The answer is, of course, that you don't measure resources easily. But as the empirical tests of resource-based theory continue to evolve, what becomes clear is that it is possible to derive testable assertions from this theory and then to collect the data needed to test these assertions.

In 1916, Einstein believed his theory of gravity waves could never be tested. In 1995, Godfrey and Hill were also not optimistic about the testability of many of the central assertions of resource-based theory. It may well be that both these predictions turn out to be overly pessimistic.

NOTES

1. When firms implement a set of incoherent, self-contradictory strategies, it is often difficult for competitors to know what exactly a firm intends to do. Of course, this is often because this firm, itself, does not know exactly what it intends to do.

2. Recall that imitation can take two forms: direct duplication or substitution (Barney, 2002). The arguments developed in this section apply most directly to direct duplication. Further work is required to see if these same arguments apply to resource substitution.

3. Ray et al. (2004) try to finesse the sustainability question by arguing that the North American insurance industry is a very mature industry and that the relationship between IT and customer service is well known in the industry. In such a setting, any remaining heterogeneity in the application of IT to the customer-service function must be the result of costly to imitate resources and capabilities possessed by some firms but not others. However, because they have only cross-sectional data, they obviously are unable to test this hypothesis directly.

REFERENCES

Amit, R., & Schoemaker, P. J. H. (1993). Strategic assets and organizational rent. *Strategic Management Journal*, *14*, 33–46.

Armour, H. O., & Teece, D. J. (1978). Organizational structure and economic performance: A test of multidivisional hypothesis. *Bell Journal of Economics*, *9*, 106–122.

Barnett, W. P., Greve, H. R., & Park, D. Y. (1994). An evolutionary model of organizational performance. *Strategic Management Journal*, *15*, 11–28.

Barney, J. B. (1986a). Strategic factor markets: Expectations, luck, and business strategy. *Management Science*, *32*, 1231–1241.

Barney, J. B. (1986b). Organizational culture: Can it be a source of sustained competitive advantage? *Academy of Management Review*, *11*, 656–665.

Barney, J. B. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, *17*, 99–120.

Barney, J. B. (2002). *Gaining and sustaining competitive advantage* (2nd ed.). Upper Saddle River, NJ: Prentice-Hall.

Barney, J. B., & Arikan, A. M. (2001). The resource-based view: Origins and implications. In: M. A. Hitt, R. E. Freeman & J. S. Harrison (Eds), *The Blackwell handbook of strategic management* (pp. 124–188). Malden, MA: Blackwell Publishers Inc.

Combs, J., & Ketchen, D. (1999). Explaining interfirm cooperation and performance: Toward a reconciliation of predictions from the resource-based view and organizational economics. *Strategic Management Journal*, *20*(9), 867–888.

Dierickx, I., & Cool, K. (1989). Asset stock accumulation and the sustainability of competitive advantage. *Management Science*, *35*, 1504–1511.

Dutta, S., Narasimhan, O., & Rajiv, S. (2005). Conceptualizing and measuring capabilities: Methodology and empirical application. *Strategic Management Journal*, *26*, 277–285.

Godfrey, P., & Hill, C. W. L. (1995). The problem of unobservables in strategic management research. *Strategic Management Journal*, *16*, 519–533.

Hansen, M. H., Perry, L. T., & Reese, C. S. (2004). A Bayesian operationalization of the resource-based view. *Strategic Management Journal*, *25*, 1279–1295.

Hatch, N. W., & Dyer, J. H. (2004). Human capital and learning as a source of competitive advantage. *Strategic Management Journal*, *25*, 1155–1178.

Henderson, R. M., & Cockburn, I. (1994). Measuring competence? Exploring firm effects in pharmaceutical research. *Strategic Management Journal*, *15*, 63–84.

Hult, G. T., & Ketchen, D. (2001). Does market orientation matter?: A test of the relationship between positional advantage and performance. *Strategic Management Journal*, *22*(9), 899–906.

Lafferty, M. (2005). *Home PCs may prove point for Einstein*. Columbus Dispatch, Tuesday, February 22, 2005, pp. A1 +.

Leiblein, M. J., & Miller, D. J. (2003). An empirical examination of transaction- and firm-level influences on the vertical boundaries of the firm. *Strategic Management Journal*, *24*, 839–859.

Lippman, S. A., & Rumelt, R. P. (1982). Uncertain imitability: An analysis of interfirm differences in efficiency under competition. *Bell Journal of Economics*, *13*, 418–453.

Makadok, R. (1999). Interfirm differences in scale economies and the evolution of market shares. *Strategic Management Journal*, *20*, 935–952.

Mansfield, E., Schwartz, M., & Wagner, S. (1981). Imitation costs and patents: An empirical study. *The Economics Journal*, *91*, 907–918.

Miller, D., & Shamsie, J. (1996). The resource-based view of the firm in two environments: The Hollywood film studios from 1936 to 1965. *Academy of Management Journal*, *39*(3), 519–543.

Nelson, R. R., & Winter, S. G. (1982). *An evolutionary theory of economic change*. Cambridge: Harvard University Press.

Porter, M. E. (1980). *Competitive strategy: Techniques for analyzing industries and competitors*. New York, NY: The Free Press.

Porter, M. E. (1991). Towards a dynamic theory of strategy. *Strategic Management Journal*, *12*(Winter Special Issue), 95–117.

Ray, G., Barney, J. B., & Muhanna, W. A. (2004). Capabilities, business processes, and competitive advantage: Choosing the dependent variable in empirical tests of the resource-based view. *Strategic Management Journal*, *25*, 23–37.

Rouse, M., & Daellenbach, U. (1999). Rethinking research methods for the resource-based perspective: Isolating sources of competitive advantage. *Strategic Management Journal*, *20*, 487–494.

Song, M., Droge, C., Hanvanich, S., & Calantone, R. (2005). Marketing and technology resource complementarity: An analysis of their interaction effect in two environmental contexts. *Strategic Management Journal*, *26*, 259–276.

Teece, D. J. (1980). Economies of scope and the scope of the enterprise. *Journal of Economic Behavior and Organization*, *1*, 223–247.

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, *18*, 509–533.

MECHANISMS AND EMPIRICAL RESEARCH

Philip Bromiley and Scott Johnson

ABSTRACT

Good research goes beyond testing the aggregate predictions of a theory to test the theory's underlying mechanism. A mechanism is a plausible account of the process that causes a systematic relationship between variables. Strategy researchers particularly need to understand the mechanisms that drive firm behavior and outcomes because we seek both to explain and offer prescriptions. We recommend that theories clearly specify their mechanisms and that empirical research test such mechanisms. Such tests will help differentiate among theories with similar aggregate predictions.

INTRODUCTION

Why do strategy scholars do empirical research? Most scholars would say that they do empirical research to test theories. We will argue that they mean or should mean they want to test the underlying *explanations*. In testing an explanation, we want to do more than simply see if the theory's primary predictions fit the data. We want to understand if the mechanism the theory postulates actually operates in the empirical world. We generally want

Research Methodology in Strategy and Management
Research Methodology in Strategy and Management, Volume 2, 15–29

ISSN: 1479-8387/doi:10.1016/S1479-8387(05)02002-3

causal explanations; we want to understand the mechanisms creating the observed phenomena.

Salmon (1998) argues that scientific explanations can take two general forms. The first form appeals to a general law. For example, early work in diversification looked for the "optimal diversification level" – a general conclusion that particular patterns of diversification associate with higher performance than other patterns of diversification.

The second form of explanation emphasizes mechanism. Salmon argues that, "the aim of explanations of this sort is to exhibit the ways in which nature operates; it is an effort to lay bare the mechanisms that underlie the phenomena we observe and wish to explain" (Salmon, 1998, p. 71). Although empirical relations may exist without explanation (e.g. Salmon offers the algebraic relation between pressure, volume, and temperature of a gas), social scientists generally want to understand why these relations hold – to explain the relations (in the gas example, this requires a deeper level of atomic theory). In the diversification example, a mechanism explanation postulates and allows testing of specific process (e.g., sharing facilities, managerial-skill transfers, etc.) that creates associations between diversification patterns and firm performance.

Although ultimately the two forms of explanation might converge, they offer very different practical approaches to research.

Even when strategy scholars carefully avoid causal hypotheses and only hypothesize associations, they almost universally interpret their results as if they had tested hypotheses reflecting causal mechanisms. That is, we might test whether "related-diversified firms have higher performance than unrelated," but we really want to know (and interpret the results as) whether "moving to related diversification from unrelated positively influences firm performance" or similar statements.

Social sciences frequently offer explanations based on mechanisms. An explanation consists of prior conditions and mechanisms or laws that operate on those prior conditions to generate predictions of new situations (Simon, 1992); we explain X by finding prior conditions A and B, then applying a theory that relates A and B to X. Thus, we predict the behavior of a firm in time $t+1$ by combining data on the firm's state at time t with a mechanism that indicates how the firm behaves given those data.

> Assume that we have observed a systematic relationship between two entities, say I and O. In order to explain the relationship between them we search for a mechanism, M, which is such that on the occurrence of the cause or input, I, it generates the effect or outcome, O. The search for mechanism means that we are not satisfied with merely establishing systematic covariation between variables or events; a satisfactory explanation

> requires that we also be able to specify the social "cogs and wheels" (Elster 1989, p. 2) that have brought the relationship into existence … a mechanism can be seen as a systematic set of statements that provide a plausible account of how *I* and *O* are linked to one another… The approach advocated here does not rest with describing the strength and the form of a relationship between entities of interest but addresses a further and deeper problem: how (i.e. through what process) was the relationship brought about?
>
> Hedström and Swedberg (1998, p. 7–10) address the need for mechanisms in sociology

Good explanations depend on features of the preconditions, mechanisms, and predictions. First, the preconditions must fit the facts. Both scientifically and in our own lives, we reject explanations based on obviously incorrect premises. For example, most would reject out of hand our children's explanations if we know they depend on untrue circumstances. If a son said he spent the afternoon in the art museum, his parents would reject this explanation if they knew the art museum was closed (assuming that the museum being open is a necessary precondition to spending time there). Second, the mechanism needs some generality – as scholars, we cannot work with idiosyncratic mechanisms for each event. Such explanations cannot build cumulative understanding. Third, when combined with the preconditions, the mechanisms should make falsifiable predictions. These predictions can include both the final topics of interest and a variety of intermediary observable features that the mechanism says should take certain values. Mechanisms that make no predictions (e.g. "it is God's will" without an exogenous specification of what God wants) are inherently untestable and therefore not suitable for scientific analysis.

Many sciences use this form of explanation (Simon, 1992). A Newtonian physicist explains the velocity of a ball rolling down a ramp using certain facts (e.g. the height of the ramp, the angle of the ramp, etc.) and mechanisms (e.g. Newton's laws). Combining facts and mechanisms creates an explanation (and prediction) of the ball's movement. Evolution and genetics explain changes in populations using reproduction/selection mechanisms and prior populations. Astronomers explain the locations of the planets using physical laws and the prior positions of the planets. These explanations explain conditions at time t as a function of conditions at time $t-1$, etc. Note that such explanations do not necessarily have to have an answer for time $t = 0$. Explaining the rotation of the planets does not require a full theory of their original formation. In strategy, theories of firm performance do not have to explain the existence of firms.

The quality of an explanation depends on the correctness of the preconditions, the generality of the mechanisms, and the accuracy of the

predictions (including those related to the mechanism *per se*). A good empirical test of an explanation should test all three as directly as possible.

The need to theorize about and empirically test mechanisms in scientific theories has appeared in numerous different contexts in several different scientific disciplines. For example, in the last half of the 19th century, German physicists were among the most advanced in the world. However, a classical view of physics that rejected using the concept of atoms to explain physical phenomena dominated German physics. Philosophers such as Mach argued that physics should only deal with quantitative empirical relations among readily observable entities. Indeed, because atoms were unobservable, Mach argued that theorizing about atoms was inherently unscientific. This position retarded the development of deeper levels of explanation and slowed the progress of German physicists. The history of physics over the last century can be seen as the development and testing of ever deeper levels of mechanisms to explain higher-level phenomena: atoms to explain the behavior of gases; electrons, protons, and neutrons, to explain atoms, and so forth.

"A good explanation" implies use of the correct mechanism. Theories in physics that predicted some outcomes quite well were rejected when the ancillary implications of the mechanisms did not hold up. Although the theory that assumed the sun and stars rotate around the earth predicted almost all of what the average individual observes in the sky, this explanation was rejected because the mechanism is incorrect.

EXPLANATIONS AND MECHANISMS IN STRATEGY RESEARCH

Bromiley (2004) suggests that strategic management research has three primary objectives:

- Explaining firm behavior at the strategic level.
- Explaining performance differences among firms.
- Providing suggestions to improve firm performance.

Strategic scholars want to understand why firms make the strategic choices they do, and how these choices and other factors interact to influence firm performance. Understanding these may let us suggest ways firms can improve performance.

Some might ask why we need to explain firm behavior to understand what determines firm performance. Simply, differences in firm behavior largely

determine firm performance. If we wish to explain someone winning a game, say tennis, part of the explanation will deal with the individual's behavior. Although we could make the explanation strictly a function of competition by assuming individuals play optimally, this violates our understanding of real behavior. Facing the same situation, individuals and firms will differ in how they want to respond. Furthermore, their actual response frequently differs from their desired response. Many software companies want to invent the next killer app, but few do. To understand who wins, we need to understand the competitors' behaviors.[1]

Simon (1946) argued that many management scholars were searching for simple rules, for example, how many subordinates a superior should supervise. He referred to these as proverbs of administration. However, if the appropriate relationships depend on the other variables, such proverbs or simple relations can be quite misleading. For example, the number of subordinates an individual can effectively supervise clearly depends on a variety of factors. Thus, instead of looking for simple relations, Simon (1946) argued scholars should attempt to understand the process by which organizations operate.

The issue of testing mechanisms instead of aggregate predictions has appeared in several areas of strategic management research. Let us consider several examples related to mechanisms in strategy research.

The Example of Top-Management Team (TMT) Literature

Due to ready availability, early research on top management teams (TMTs) used demographic variables such as top manager functional backgrounds, education, sex, age, and tenure with the firm to explain performance. The theorizing used group process or information processing mechanisms. For example, researchers would postulate that cognitive diversity helps the team understand and thus adapt to changing environments. Empirically, they then used demographic diversity to proxy for cognitive diversity.

However, diversity could affect performance through multiple mechanisms; we have multiple theories that predict heterogeneous management teams will be more effective than homogeneous. An information-based explanation argues team heterogeneity increases the amount of information a team has and thus heterogeneous teams should have more information than homogeneous teams. A conflict-based explanation argues that conflict, particularly task conflict, helps groups work through their assumptions and their logics and therefore results in better choices. A network-based

explanation would argue that heterogeneous teams have access to the resources of larger networks and using these networks improves performance. A cognitive-based explanation would argue that demographic heterogeneity coincides with cognitive heterogeneity and cognitive heterogeneity improves the ability of the team, at least partially by avoiding groupthink and similar problems.

These distinctions matter for both scholars and practitioners. Academically, we clearly want to know which of these mechanisms actually influences outcomes. The different theories also imply differing limiting conditions – for example, heterogeneity that did not result in task conflict should not benefit the firm under the conflict explanation and neither should heterogeneity if the managers do not have differing networks in the network explanation.

Practically, each explanation suggests different tactics for improving TMT decisions. If we want to derive recommendations from findings regarding team heterogeneity and performance, we need to know which of these mechanisms operates. For example, if the explanation rests on networks, we should use direct measures of networks instead of demographic heterogeneity. If conflict generates the benefits, we should consider directly influencing or at least not reducing conflict. If information generates the benefits, then our recommendations should only apply to situations where breadth of information may have value. Prescription requires understanding the underlying mechanism.

The Example of Diversification Research

Many strategy papers have tested hypotheses about the relation between diversification and performance. Although they often find that related-diversified companies perform better than unrelated, the results are mixed, and many diversified companies perform better than less-diversified ones.

We have many theories to explain associations between diversification and performance:

1. Internal capital markets – Managers operating in internal capital markets (i.e. allocating resources within a firm) have better information and better control than outside investors making internal capital markets beneficial. This is particularly likely to hold in countries with weak public capital markets (see e.g. Lins & Servaes, 1999).
2. Economies of scope – Producing or selling multiple products may offer economies related to the breadth of operations. One example of this comes in branding where advertising for a brand name may benefit

multiple products sold under a given name. Alternatively, the products may share physical production or distribution facilities. In some areas, the waste products from producing one product may be an important input in producing the other product (see e.g. Singh & Montgomery, 1987).

3. Following the dominant logic argument, the organizational systems in any corporation and the thinking of management have commonalities across business units. If the business units differ greatly, the systems and thinking may be inappropriate thus lowering performance (see e.g. Prahalad & Bettis, 1986).
4. Resources – If a firm has special abilities that it cannot sell directly, it may extract value from those abilities by applying them to multiple businesses. Firms whose diversification uses such resources should perform better than ones whose high levels of diversification preclude the existence of specific resources helping all business units (see e.g. Chatterjee & Wernefelt, 1991).
5. Risk reduction – If capital markets or other markets value the likelihood of continued survival of a firm or stability of a firm's profit streams, then firm might lower its risk by diversification and so increase its value. If customers or suppliers value low risk partners, such lowered risk may result in performance differences (see e.g. Lubatkin & Chatterjee, 1994).
6. Control and incentives – Internal management may offer the business unit stronger (or weaker – this is an untested assumption) control and managerial incentives than a freestanding business making internalization good (or bad) for business performance (see e.g. Aron, 1988).
7. Transaction costs economics – Firms should internalize when asset specificity, uncertainty, and frequency of transactions are high. Firms doing this correctly may prosper whereas ones doing it incorrectly have lower performance (see e.g. Amit, Livnat, & Zarowin, 1989).
8. Merger activity – Low performing firms may acquire other businesses in hopes of improving corporate performance (see e.g. Lang & Stulz, 1994).

This list suggests several immediate conclusions.

First, we should not search for general relations between diversification and performance. If some of these explanations have validity, the appropriate level of diversification depend on other factors. Like the management proverbs (Simon, 1946), a general rule for diversification would misrepresent the complexity of the problem.

Second, with multiple possible explanations, an empirically observed association between relatedness and performance can tell us little about the correctness of a particular explanation.

Third, since the multiple explanations have slightly different implications and conditions, our statistical results will be weaker and less informative than they would be if we considered all the appropriate explanations. For example, branding should not apply to relatedness if those products do not have brand reputations. The physical production economies of scope should only apply to products with such relations. An aggregate relatedness measure provides a poor proxy for both of these causal mechanisms and so should not explain the behaviors as well as a measure more clearly tied to the specific construct.

Fourth, and most problematic, scholars who use a specific argument to justify a hypothesis for a relation between relatedness and performance often see that relation as supporting their argument. Thus, they may conclude support for their mechanism and not recognize that other mechanisms make the same aggregate prediction.

The Example of Transaction Cost Economics (TCE)

How strong is the empirical support for Transaction Cost Economics (TCE)? The answer to this question depends on whether one looks at the underlying mechanisms or just the aggregate predictions of the theory. Many studies use TCE to derive predictions regarding such things as make-versus-buy decisions, organizational form, strategic alliances, and acquisitions. These predictions are frequently supported, though not always (see Walker & Weber, 1984). Williamson (1999) has noted that the "number of published studies exceeds 400 and involves scientists in Europe, Japan, India, Mexico, South America, New Zealand, and the list goes on… the theory and evidence display a remarkable congruity." If one looks only at the aggregate predictions of the theory, the empirical evidence generally supports TCE.

However, many scholars question the track record of TCE research. As Simon says:

> A fundamental feature of the new institutional economics is that it retains the centrality of markets and exchanges. All phenomena are to be explained by translating them into (or deriving them from) market transactions based on negotiated contracts, for example, in which employers become 'principals' and employees become 'agents'. Although the new institutional economics is wholly compatible with and conservative of neoclassical theory, it does greatly multiply the number of auxiliary exogenous assumptions that are

> needed for the theory to work. For example, to explain the presence or absence of certain kinds of insurance contracts, moral risk is involved; the incompleteness of contracts is assumed to derive from the fact that information is incomplete or distributed asymmetrically between the parties to the contract. Since such constructs are typically introduced in the analysis in a casual way, with no empirical support except an appeal to introspection and common sense, mechanisms of these sorts have proliferated in the literature, giving it a very *ad hoc* flavor.
>
> In general, the new institutional economics has not drawn heavily from the empirical work in organizations and decision-making for its auxiliary assumptions. (Simon, 1991, pp. 26, 27)

Note the different understandings of empirical support. Simon (1991) does not deny that the general predictions of TCE roughly fit the data; asset specificity and uncertainty do correlate with whether a transaction takes place in markets or firms. However, the TCE research does not offer empirical support of the theory's internal mechanisms and assumptions. Such support could come from previous empirical results or direct tests in the context of the theory.

TCE assumes organizations must act as if individuals are self-interested and amoral. Without this assumption, the mechanism of TCE cannot operate but Simon (1997), along with Sen (1970, 1977) and others, argue that organizations of strictly self-interested individuals would not function. Instead, these authors suggest other explanations for co-operation in organizations including identification and pride in work (see Tyler, 1999).

Sen and others argue that control mechanisms cannot induce adequate cooperation from completely self-interested employees who get no pleasure from their jobs. Consider academics. We all know that schools offer little reward for mentoring, service, and collegiality; if all faculty members did only what was rewarded, we would find little service, mentoring, or collegiality. If schools use student satisfaction surveys to judge and reward teaching in pay and promotion decisions, completely self-interested instructors should design courses to maximize student satisfaction. If schools do not use such surveys, faculty should design courses to minimize effort. Everyone would use multiple choice exams rather than essays or short-answer exams. Schools have little direct control over senior faculty; if they are self-interested they should teach badly and do no service. After all, schools seldom dismiss senior faculty members for lack of service or for bad teaching, and at the end of a career, the pay incentives facing most faculty members are trivial.

Non-academics face similar issues; formal incentives can only address part of any job (see Simon, 1991). If employees really do the minimum the

rules require (termed a work to rule – a labor tactic that often largely stops production), or only what the system rewards, most organizations will cease to function. If employees were really fully self-interested and amoral, managers could do little; even firing current employees would simply mean replacing them with new employees who would be just as self-interested and amoral as the old employees.

Regarding TCE, scholars have offered two very different explanations to explain why internalization improves performance in certain circumstances. According to TCE, organizations offer more effective controls than markets for exchanges when contracts are difficult to write and enforce. According to Simon (1991, 1997), the benefits of bringing employees into an organization come partially from employees identifying with the organization, not just from controls (see Kogut & Zander, 1996).

Empirical work on TCE seldom, if ever, differentiates between these explanations – *it does not test the causal mechanism*. Instead, the studies look for associations between the difficulty of the contracting problem (asset specificity, uncertainty) and organizational arrangements. Such analysis does not address why the organizational arrangement works.[2]

For both scholars and practitioners, it makes a big difference whether internalization works by controlling self-interested, amoral employees, or by employees developing identification with the organization. Indeed, both mechanisms may operate, but empirical tests of aggregate predictions that do not differentiate between the mechanisms do not provide support for either theory.

The Example of R&D and Advertising Intensity

One of the oddest examples of the no-mechanism problem comes when scholars working with different theories use the same variable as a proxy for different constructs. They then find the variable relates to the dependent variable and conclude support for their theories.

Ancillary assumptions often relate our constructs to empirical measures. That is, particularly for secondary data, we argue that some readily available data reflect our constructs and then use those data as proxies. For example, one of the authors of this paper used variance in analyst forecasts of firm profitability as a proxy for the uncertainty of the firm's income stream (Bromiley, 1991). Many studies use industry average performance as a reference point.

Consider strategy's uses of advertising-to-sales or R&D-to-sales ratios. Scholars have used them to reflect several different constructs:

1. TCE studies use them to indicate specific assets for which contracts are difficult to write (see e.g. Mosakowski, 1991; Regan, 1997).
2. Resource-based-view scholars have seen them as implying unobservable but important resources (see e.g. Chatterjee & Wernerfelt, 1991; Dierickx & Cool, 1989).
3. Industrial organization scholars view them as barriers to entry or exit (see e.g. Hirschey, 1981; Kessides, 1986; Lustgarten & Thomadakis, 1987; Mueller & Tilton, 1969).
4. Some strategy researchers see them as indicators of product differentiation and technology strategies (see e.g. Erickson & Jacobson, 1992; Grabowski & Mueller, 1978; Zahra & Covin, 1993).
5. Other researchers use them to measure actual expenditures on sales and R&D – budgetary expenditures based on some internal management systems (see e.g. Grabowski & Baxter, 1973; Hoskisson & Hitt, 1988).

Several theories predict a positive relation between these variables and performance. How can we differentiate among these theories?

If the theories made specific quantitative predictions about the relations, we could use aggregate estimates to differentiate among the theories, but few of our theories make quantitative predictions. Our theories say "a positive association," not "the coefficient should be 5."

Aggregate analysis cannot differentiate among these causal mechanisms. We need theories that specify clear causal mechanisms and more micro-level empirical work to see which causal mechanism actually causes the aggregate relations (Bromiley, 1981). This creates problems for theories with unclear causal mechanisms.

A RESEARCH APPROACH BASED ON MECHANISMS

What does the mechanisms approach imply for the kind of research that strategic management scholars do? The mechanisms-based approach has implications for both theoretical and empirical work.

Specify the Mechanism that Drives the Theory

Scholars who build theories need to clarify the precise mechanisms of their theories. Indeed, without clear mechanisms, theories are ill-formed. Exactly how do firms come to use "optimal structures" as defined by TCE? In agency theory, exactly what mechanisms determine the governance system

in a firm? Fama and Jensen (1983) suggest foresight and calculation by the CEO and potential investors result in firms using governance structures that minimize agency costs. Jensen and Meckling (1976) propose an evolutionary mechanism where competition eliminates firms with the wrong structures. Many strategy papers associate governance (and incentive) structures with difference in firm performance, implicitly assuming firms often have incorrect governance/incentive structures. Without clear mechanisms, the theories remain incomplete.

Test the Plausibility of the Underlying Assumptions

As noted earlier, an explanation includes a description of prior conditions and the underlying mechanism that connects prior conditions to predictions. Good empirical work will examine all three – preconditions, mechanisms, and predictions.

Often, we can quickly reject proposed theories or mechanisms because they depend on grossly incorrect assumptions. Just as we would reject explanations from our children that assert clearly implausible facts, we can reject scholarly explanations that assume implausible facts. As always, some moderation is required – any theorizing involves simplification. However, some theories make extremely implausible assumptions that cannot be seriously defended as an actual mechanism.

One frequent form of implausible assumption comes in assuming managers know things they do not or make calculations they cannot. This reflects the bounded rationality critique of rational models, where rational model advocates frequently assume managers know things that managers clearly do not know (Bromiley, 2004; Simon, 1947).

Design Research that Tests Mechanisms

Just as we test final predictions, we can test the implications of differing mechanisms. Indeed, we find scholarship often starts with general tests of high-level association and then moves to clearer theories and tests of the underlying mechanisms.

For example, studies that analyze information processing within TMTs give instructive examples of uncovering the underlying mechanism behind a phenomenon. As noted above, early work used cognitive or group process theories to associate TMT demographics with outcomes. More recent work

directly examines such processes. Schweiger, Sandberg, and Rechner (1989) used a laboratory experiment that manipulated management team decision processes and found that differences in decision process influenced decision quality. Amason (1996) directly measures TMT cognitive and affective conflict to assess the influence of types of conflict on decision quality. Simons, Pelled, and Smith (1999) examine how trust among TMT members changes the impact of conflict on performance. Rau (2001) studies the impact of TMT's knowledge distribution and knowledge sharing on the performance of banks. All of these studies go beyond predictions of how aggregate measures of TMT characteristics affect the firm to uncover the processes that relate TMT characteristics to organizational outcomes.

As scholars continue to examine mechanisms more closely, they will find that interviews with the individuals making the decisions of interest can inform both theorizing and tests of mechanisms. Such interviews can tell us what managers do and do not know, what managers think they do, and, often, what factors managers consider in making such decisions. Although such interview data cannot stand as a final test, interviews offer important information for theory building and testing.

As strategic management scholars, we want to understand what influences firm behavior and performance. Such understanding requires clear specification and testing of the mechanisms underlying our aggregate theories. Such understanding also opens the door to richer and more accurate theories and more legitimate testing of theories.

NOTES

1. See Bromiley (2004) for a more general discussion of this issue.
2. TCE also implicitly includes another problematic causal mechanism – by assuming firms use optimal governance structures, it assumes firms know such structures. However, the theory offers no plausible mechanism by which firms would learn the optimal structures. Again, a clear examination of mechanisms could enrich the area.

REFERENCES

Amason, C. (1996). Distinguishing the effects of functional and dysfunctional conflict on strategic decision making: Resolving a paradox for top management teams. *Academy of Management Journal*, *39*(1), 123–148.

Amit, R., Livnat, J., & Zarowin, P. (1989). The mode of corporate diversification: Internal ventures versus acquisitions. *Managerial and Decision Economics*, *10*(2), 89–100.

Aron, D. J. (1988). Ability, moral hazard, firm size, and diversification. *The Rand Journal of Economics*, *19*(1), 72–87.

Bromiley, P. (1981). Task environments and budgetary decision-making. *Academy of Management Review*, *6*(2), 277–288.

Bromiley, P. (1991). Testing a causal model of corporate risk-taking and performance. *Academy of Management Journal*, *34*(1), 37–59.

Bromiley, P. (2004). *The behavioral foundations of strategic management*. Oxford: Blackwell.

Chatterjee, S., & Wernerfelt, B. (1991). The link between resources and type of diversification: Theory and evidence. *Strategic Management Journal*, *12*(1), 33–48.

Dierickx, I., & Cool, K. (1989). Asset stock accumulation and sustainability of competitive advantage. *Management Science*, *35*(12), 1504–1511.

Erickson, G., & Jacobson, R. (1992). Gaining comparative advantage through discretionary expenditures: The returns to R&D and advertising. *Management Science*, *38*(9), 1264–1279.

Fama, E. F., & Jensen, M. C. (1983). Separation of ownership and control. *Journal of Law and Economics*, *26*(2), 301–325.

Grabowski, H. G., & Baxter, N. D. (1973). Rivalry in industrial research and development: An empirical study. *The Journal of Industrial Economics*, *21*(3), 209–235.

Grabowski, H. G., & Mueller, D. C. (1978). Industrial research and development, intangible capital stocks, and firm profit rates. *The Bell Journal of Economics*, *9*(2), 328–343.

Hedström, P., & Swedberg, R. (1998). Social mechanisms: An introductory essay. In: P. Hedström & R. Swedberg (Eds), *Social mechanisms: An analytical approach to social theory*. New York: Cambridge University Press.

Hirschey, M. (1981). The effect of advertising on industrial mobility, 1947–1972. *The Journal of Business*, *54*(2), 329–339.

Hoskisson, R. E., & Hitt, M. A. (1988). Strategic control systems and relative R&D investment in large multiproduct firms. *Strategic Management Journal*, *9*(6), 605–621.

Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, *3*, 305–360.

Kessides, I. N. (1986). Advertising, sunk costs, and barriers to entry. *The Review of Economics and Statistics*, *68*(1), 84–95.

Kogut, B., & Zander, U. (1996). What do firms do? Coordination, identity, and learning. *Organization Science*, *7*(5), 502–518.

Lang, L. H. P., & Stulz, R. M. (1994). Tobin's Q, corporate diversification, and firm performance. *Journal of Political Economy*, *102*(6), 1248–1280.

Lins, K., & Servaes, H. (1999). International evidence on the value of corporate diversification. *Journal of Finance*, *54*(6), 2215–2239.

Lubatkin, M., & Chatterjee, S. (1994). Extending modern portfolio theory into the domain of corporate diversification: Does it apply? *Academy of Management Journal*, *37*(1), 109–136.

Lustgarten, S., & Thomadakis, S. (1987). Mobility barriers and Tobin's Q. *The Journal of Business*, *60*(4), 519–537.

Mosakowski, E. (1991). Organizational boundaries and economic performance: An empirical study of entrepreneurial computer firms. *Strategic Management Journal*, *12*(2), 115–133.

Mueller, D. C., & Tilton, J. E. (1969). Research and development costs as a barrier to entry. *The Canadian Journal of Economics*, *2*(4), 570–579.

Prahalad, C. K., & Bettis, R. A. (1986). The dominant logic: A new linkage between diversity and performance. *Strategic Management Journal*, *7*(6), 485–501.

Rau, D. (2001). *Knowing who knows what: The effect of transactive memory on the expertise, diversity-decision quality relationship in managerial teams*. Unpublished doctoral dissertation, University of Minnesota, Carlson School of Management, Minneapolis.

Regan, L. (1997). Vertical integration in the property-liability insurance industry: A transaction cost approach. *The Journal of Risk and Insurance*, *64*(1), 41–62.

Salmon, W. C. (1998). *Causality and explanation*. Oxford: Oxford University Press.

Schwiger, D. M., Sandberg, W. R., & Rechner, P. L. (1989). Experiential effects of dialectical inquiry, devil's advocacy, and consensus approaches to strategic decision making. *Academy of Management Journal*, *32*(4), 745–772.

Sen, A. K. (1970). *Collective choice and social welfare*. San Francisco: Holden-Day.

Sen, A. K. (1977). Rational fools: A critique of the behavioral foundations of economic theory. *Philosophy and Public Affairs*, *6*(4), 317–344.

Simon, H. A. (1946). The proverbs of administration. *Public Administration Review*, *6*, 53–67.

Simon, H. A. (1947). *Administrative Behavior*. New York: Free Press.

Simon, H. A. (1991). Organizations and markets. *Journal of Economic Perspectives*, *5*, 25–44.

Simon, H. A. (1992). What is an 'explanation' of behavior? *Psychological Science*, *2*, 150–161.

Simon, H. A. (1997). *An empirically based microeconomics*. New York: Cambridge University Press.

Simons, T., Pelled, L. H., & Smith, K. A. (1999). Making use of difference: Diversity, debate, and decision comprehensiveness in top management teams. *Academy of Management Journal*, *42*(6), 662–673.

Singh, H., & Montgomery, C. A. (1987). Corporate acquisition strategies and economic performance. *Strategic Management Journal*, *8*(4), 377–386.

Tyler, T. R. (1999). Why people cooperate with organizations: An identity-base perspective. *Research in Organizational Behavior*, *21*, 201–246.

Walker, G., & Weber, D. (1984). A transaction cost approach to make-or-buy decisions. *Administrative Science Quarterly*, *29*(3), 373–391.

Williamson, O. E. (1999). Strategy research: Governance and competence perspectives. *Strategic Management Journal*, *20*(12), 1087–1108.

Zahra, S. A., & Covin, J. G. (1993). Business strategy, technology policy and firm performance. *Strategic Management Journal*, *14*(6), 451–478.

STRATEGIC MANAGEMENT STUDIES ARE A SPECIAL CASE FOR META-ANALYSIS

Dan R. Dalton and Catherine M. Dalton

ABSTRACT

Meta-analysis has been relied on relatively infrequently in strategic management studies, certainly as compared to other fields such as the medical sciences, psychology, and education. This may be unfortunate, as there are several aspects of the manner in which strategic management studies are typically conducted that make them especially appropriate for this approach. To this end, we provide a brief foundation for meta-analysis, an example of meta-analysis, and a discussion of those elements that strongly recommend the efficacy of meta-analysis for the synthesis of strategic management studies.

A BRIEF HISTORY OF RESEARCH SYNTHESIS

The development of meta-analytical approaches was a proposed solution to the general need for research synthesis. For decades prior to the introduction of meta-analysis, it was well known that individual studies – however well executed – would be unlikely to settle whatever research question was at

Research Methodology in Strategy and Management
Research Methodology in Strategy and Management, Volume 2, 31–63

ISSN: 1479-8387/doi:10.1016/S1479-8387(05)02003-5

issue in a given field of inquiry. Moreover, it was recognized that individual studies, replications, and constructive replications (e.g., Lykken, 1968) conducted in similar areas of inquiry could produce highly variable results. Differences, for example, in samples, measures, and protocols would always introduce some variability in the relationship of interest (for general discussion, see, e.g., Hunter, Schmidt, & Jackson, 1982; Hunter & Schmidt, 2004; Rosenthal & DiMatteo, 2001; Rothstein, McDaniel, & Borenstein, 2002).[1] Accordingly, an enduring challenge for researchers has been to identify a reliable means to combine the data/results from multiple studies to yield some summary conclusion.

The earliest example of a formal approach to synthesizing the results of multiple studies was provided by Karl Pearson (1904), in which he averaged correlations from five studies that examined the relationship between inoculation for typhoid fever and mortality. Others, too, attempted to combine effect sizes (e.g., Cochran, 1937, 1943; Pearson, 1933; Thorndike, 1933; Yates & Cochran, 1938), while others followed and sought to combine the significance levels across multiple studies (e.g., Fisher, 1932; Pearson, 1938; Tippett, 1931). Despite these early efforts at synthesizing results across studies, there was apparently no enthusiasm for a quantitative approach to research synthesis until 1975 or so. In those intervening years, the approach of choice was commonly referred to as a "narrative review."

The Narrative Review

The typical approach for a narrative review is to identify relevant studies – we will develop 10 studies as examples to be followed – that include the variables of interest and to assemble a table of results. Invariably, the table will reflect some studies with a statistically significant positive relationship, some negative, and some that did not reach the level of statistical significance. In most cases, the results are synthesized through a "count." A summary, for example, might note that the majority of the studies (6) were positive, two were negative, and two were not significant. Accordingly, the literature review might conclude that the preponderance of evidence supports a positive relationship between the variables of interest. More often, though, most reviews of this type would conclude that the data are inconsistent and allow no strong consensus about the nature of the relationship between the independent and dependent measures of interest.

Two examples of this approach in the strategic management studies literature may underscore these issues.[2] The first is a review of several elements

of organizational structure and performance (Dalton, Todor, Fielding, Spendolini, & Porter, 1980). The data on which this research relied is, in retrospect, a rather simple matrix noting the variables, the nature of the sample, and the direction of the reported relationship (+, 0, −). Neither the actual effect sizes nor the sample size were noted. Beyond that, there was little apparent concern about the vastly different measures of performance that were apparently considered in concert. The summary of these "analyses" concluded that "The literature on structure–performance relationships is among the most vexing and ambiguous in the field of management … the nature and direction of these relationships are tenuous. Our review has underscored the relative lack of generalizability of research in the area …" (Dalton et al., 1980, p. 60).

Another narrative review examined the relationship between strategic planning and organizational performance (Shrader, Taylor, & Dalton, 1984). This research, too, relied on a matrix of planning variables, the nature of the sample, performance measures (15 types), and the direction of the reported relationships (+, 0, −). With no attempt to link the performance measures, no indicators of sample size, or actual effect sizes, the overall conclusion noted that "the existing literature … allowed limited conclusions and those offered were made with some reticence" (Shrader et al., 1984, p. 163). It was narrative reviews like these, and, of course, hundreds of other examples across many disciplines preceding them that propelled a resurgence of interest in some methods of synthesizing research that might overcome much of the equivocation so often seen in the review of extant literatures (e.g., Cooper, 1998; Hunter & Schmidt, 2004; Lipsey & Wilson, 2001; Rosenthal & DiMatteo, 2001).

TOWARD META-ANALYSIS

Perhaps, a provocative narrative review may have inspired meta-analysis. Hans Eysenck (1952) reported, based on literally hundreds of studies with positive, null, and negative results, that psychotherapy was of no benefit to clinical patients. As might be expected, this conclusion was highly controversial. Nearly 25 years after Eysenck's observation, Gene Glass (1976) published the first article proposing a new approach for synthesizing research, an approach to which he referred as "meta-analysis." The following year, he and his colleague Mary Lee Smith reviewed 375 relevant studies and reported, relying on the newly developed meta-analysis approach, that psychotherapy did, in fact, have therapeutic value (for other

interesting examples of path breaking meta-analysis in the medical sciences see Hunt, 1997).

Meta-analysis was apparently not universally appreciated at that time. Hans Eysenck (1978, p. 517), for example, referred to this new method as "an exercise in mega-silliness." Indeed, even as late as 1995, meta-analysis still had its critics. Sohn (1995, p. 209; see also, Meehl, 1990), for example, observed that "Meta-analytic writers have created the impression, with a farcical portrayal of the scientific process, that the process of arriving at truth is mediated by a literature review… after some critical mass of findings have been gathered, someone decides to see what all of the findings mean … and thereby knowledge is established."

Such perspectives, however, have not prevailed. With the pioneering, and largely parallel, work of Glass and his colleagues (e.g., Glass, 1977; Glass, McGraw, & Smith, 1981; Smith & Glass, 1977; Smith, Glass, & Miller, 1980), the work of Hunter and Schmidt (Hunter & Schmidt, 1990, 2004; Hunter, Schmidt, & Jackson, 1982; Schmidt & Hunter, 1977), and contributions by others (e.g., Hedges, 1982; Hedges & Olkin, 1985; Rosenthal, 1978, 1984, 1991; Rosenthal & Rubin, 1978, 1982; see also, Appendix A for a compendium of the major books and summary materials for meta-analysis), the use of meta-analysis has increased at an extraordinary pace (see Table 1). As illustrated in Table 1, we see rather obvious differences in the use of meta-analysis across disciplines, with psychology and medical sciences having the most applications and strategic management studies the fewest (see Appendix B for a full listing of citations relying on meta-analysis in strategic management studies). This pattern suggests that there is an opportunity to leverage this methodological approach, as meta-analysis is especially well suited for synthesizing strategic management studies research. Perhaps this observation is best illustrated with an example.

An Example of Meta-Analysis

Suppose we were interested in knowing whether the extent of equity holdings by CEOs in publicly traded firms is related to firms' financial performance. And, suppose that we have decided that the appropriate measure of financial performance is return on common stock, and appropriately market-adjusted. Furthermore, assume that the literature suggests that this relationship may be informed by the "maturity" of the firm, which we operationalize as a dichotomous variable with one category representing firms that are five or less years post-IPO, and the second category including all other firms.

Table 1. Growth of Meta-Analysis Applications[a].

Psychology (PsycINFO Info Data Base)	
Years	Number of Meta-Analyses
1975–1985	221
1986–1995	825
1996–2004	1,264
	Total = 2,310
Medical Science (Medline Data Base)	
1975–1985	53
1986–1995	1,271
1996–2004	3,608
	Total = 4,932
Education (Eric Data Base)	
1975–1985	164
1986–1995	268
1996–2004	209
	Total = 641
Strategic Management Studies (ABI/Inform & EBSCO)	
1975–1985	1
1986–1995	6
1996–2004	14
	Total = 21

[a]These numbers are understated. Many studies do not use the expression "meta-analysis" in the title, but do, in fact, rely on meta-analysis.

Table 2 provides a hypothetical example of the input data that would be necessary for meta-analysis to examine the assumptions. Note that the two left-most columns are not input data and are not required for meta-analysis input. Rather, this information is included to facilitate discussion in the subsequent sections.

From the onset, it is immediately apparent that a narrative review of these studies would present a challenge. Notice that there are 30 studies, 10 of which are not statistically significant, 10 of which are positive (and significant), and 10 of which reflect negative relationships (also statistically significant). "Counting" these, as previously described, would provide a very poor summary of this work. A brief review of each column of required data may establish why meta-analysis is very well suited for strategic management studies research. The first column of note in Table 2 is r, the simple bivariate correlation of the variables of interest.

Table 2. Input Data for Meta-Analysis[a].

ID	$p \leqslant 0.05$	r	n	Reliability of Y	Reliability of X	Range Restriction of Y	Range Restriction of X	Moderator
1	n	0.26	56	0.8	0.8	1	1	1
2	y	0.39	225	0.8	0.8	1	1	1
3	y	0.37	192	0.8	0.8	1	1	1
4	y	0.29	146	0.8	0.8	1	1	1
5	n	0.23	70	0.8	0.8	1	1	1
6	y	0.28	288	0.8	0.8	1	1	1
7	y	0.41	392	0.8	0.8	1	1	1
8	y	0.49	182	0.8	0.8	1	1	1
9	n	0.31	38	0.8	0.8	1	1	1
10	y	0.11	325	0.8	0.8	1	1	2
11	y	0.08	622	0.8	0.8	1	1	2
12	n	0.38	26	0.8	0.8	1	1	1
13	y	−0.28	1125	0.8	0.8	1	1	2
14	y	−0.09	484	0.8	0.8	1	1	2
15	n	0.21	80	0.8	0.8	1	1	1
16	n	0.30	40	0.8	0.8	1	1	1
17	y	−0.31	822	0.8	0.8	1	1	2
18	y	−0.22	196	0.8	0.8	1	1	2
19	n	0.21	84	0.8	0.8	1	1	1
20	y	−0.12	311	0.8	0.8	1	1	2
21	y	−0.14	214	0.8	0.8	1	1	2
22	n	0.17	120	0.8	0.8	1	1	1
23	y	−0.17	440	0.8	0.8	1	1	2
24	y	−0.23	220	0.8	0.8	1	1	2
25	y	−0.13	228	0.8	0.8	1	1	2
26	n	0.29	45	0.8	0.8	1	1	1
27	n	0.28	48	0.8	0.8	1	1	1
28	y	0.06	1088	0.8	0.8	1	1	2
29	y	−0.27	776	0.8	0.8	1	1	2
30	y	0.07	802	0.8	0.8	1	1	2

[a]The two left-most columns are not part of the input data for meta-analysis. The information in those columns is included only for reference in the text portions of the chapter.

r – A Bivariate Correlation

The early development and applications of meta-analysis were in experimental psychology, industrial–organizational psychology, organizational behavior, medical sciences, and education. Much of the body of research, at that time, did not rely on regression analyses. Accordingly, much of the

early literature will allude to an effect size referred to as *d*. This is an effect size that is the standardized difference in the means of two groups of some variable.

This is not reflective of the tradition in strategic management studies, where the use of regression analyses was, and is, far more common. We are not familiar with a meta-analysis in strategic management studies that does not rely on *r*, the simple bivariate correlation as an effect size. There are three issues of some importance here. First, the variables from which the *r* is reported need not have been the main focus of a given study to be included in a meta-analyses. It could have been a control variable with virtually no contribution to the original analyses or the study. It is only necessary that a simple correlation (*r*) between the variables of interest be available for the relationship(s) to be included in a meta-analysis.

Second, the researcher will not want to rely on r^2. The algorithms for meta-analysis software would ignore it in any case and interpret the r^2 entry as an *r*. Beyond that, the use of r^2 would disguise the directionality of the data (i.e., an *r* of 0.4 and an *r* of -0.4 are both $r^2 = 0.16$).

Third, occasionally, the researcher will identify a relevant study (i.e., it relies on the appropriate variables), but the data do not provide an *r*. This most often happens when an article does not include a correlation matrix (more about this in a subsequent section). Furthermore, the research report may not rely on regression analyses. Fortunately, it is possible to derive an *r* from virtually any research report. If the article includes an analysis relying on a *d*, *t*, *F*-score, *z*, or χ^2, these scores can be easily converted to an *r*. More esoteric conversions are also available as options (e.g., Lipsey & Wilson, 2001; Rosenberg, Adams, & Gurevitch, 2000; Rosenthal & DiMatteo, 2001).

An r Tip or Two

Sometimes, especially in older research, it will be reported that a relationship is "not statistically significant." The paper probably does note the sample size, but beyond that, there are no other results reported. In such cases *r*, or any of the other summary scores on which we normally rely, will not be available. The issue, then, is what to enter into the *r* column for the meta-analysis. In such a case, however, there is often enough information to impute an *r* value. The equation for *r* is known ($r = Z/\sqrt{n}$). Suppose that we know that $n = 120$ and $Z = 1.96$ (i.e., we would like to know the value of *r* if it was just significant at the 0.05 level). With the equation, we know that an

r of 0.179 would have been just significant at the 0.05 level. Given this, we know that the "statistically insignificant" result that the paper reported could have been anything between approximately 0.17 and −0.17.[3]

While we recognize that this approach is conservative (e.g., Lipsey & Wilson, 2001; for an extensive discussion of this point, see Pigott, 1994), we are strongly inclined to enter zero, as it is the best estimate. To do otherwise, especially for studies with small sample sizes, is potentially hazardous. Suppose, for example, the sample size is 30. By the same formula, we know that an r of 0.358 is just significant at the 0.05 level. To enter 0.35 (just under statistical significance) instead of zero would be a bit too aggressive for our comfort.

Remember, however, that with many studies included in the meta-analysis, and only a few with the zero/other choice, the results are unlikely to be substantially affected. With smaller numbers of studies, however, these decisions could be consequential. Minimally, a good guideline for all of us is to carefully report our approach for handling such cases and to be consistent with our approach. Any reliance by the same researchers on the "zero" approach in one meta-analysis and on something other than this in another meta-analysis would require a compelling rationale.

There is another suggested approach: Do not include the study if there are incomplete data. But for us, this is not a comfortable choice. We would never ignore these data or exclude them. Suppose the meta-analysis comprised of 30 studies, four of which have the zero/other issue. A comparison of the meta-analysis solution with full data set with (whatever decision the researcher has made about zero/other) the meta-analysis solution with 26 studies (with the four studies in question temporarily deleted) can easily be done. Once again, the issue is clarity in reporting these differences, if any, and how the results inform the choices for selecting the final meta-analysis data set.

Another technical point warrants mentioning. It is known that the sample correlation, r, is not an unbiased estimator of the population r (e.g., Hunter & Schmidt, 2004). Accordingly, there are those who suggest that effect sizes relied on for meta-analysis should be transformed (e.g., Rosenberg et al., 2000). Hunter and Schmidt (2004) remind us of two issues that are relevant to this discussion. First, with sample sizes of 20 or more, the bias in the sample r is "less than rounding error" (Hunter & Schmidt, 2004, p. 56). Furthermore, relying on a Fisher z-transformation (the transformation of choice for r) is often not effective. The positive bias introduced by the transformation is larger than the negative bias in the unaltered r. Hunter and Schmidt (2004, p. 56) suggest that "It is always less accurate to use the

z-transformation." And, Schulze (2004, p. 193) suggests that "differences in bias favor *r* over *z* but are minuscule in absolute value." Accordingly, we would not transform the *r* for the meta-analysis input data.

Another aspect of *r* in a meta-analysis is distinctive. Its algorithms do not rely on the notion of statistical significance. Whether a given correlation is statistically significant is of no consequence. Moreover, it is certainly not assumed that such an *r* is a null finding of the order of "there is no evidence of a relationship." However, in a subsequent section, we note that the statistical significance of an *r* to be used in a meta-analysis is not even entered in the input data.

Non-reliance on statistical significance is important. Meta-analysis does not use statistical significance of input data as any measure of its value. Moreover, it is interesting that a series of *r*'s in the same direction across a number of studies can be a far more powerful influence in the accumulation of studies than a single significant result. Rosenthal and DiMatteo (2001; see also, Cohn & Becker, 2003) provide fascinating examples of this phenomenon. Two results at a 0.06 level of statistical significance are of a much stronger evidence ($p \leqslant 0.014$) of a relationship than a single study at 0.05. Moreover, consider that 10 results at the 0.10 level of statistical significance are a much stronger evidence for a relationship ($p \leqslant 0.000025$) than five studies at the 0.05 level ($p \leqslant 0.00012$). Among a host of other reasons, this is a crucial rationale for recommending meta-analysis over the narrative review.

The n Column

The *n* column in Table 2 is simply the sample size from which the *r* was calculated. Since the firm is the level of analysis, we know from entry ID #1 that the *r* is 0.26 and this was derived from 56 companies. Meta-analytic procedures require that each observed correlation be weighted by the study's sample size in order to calculate the mean weighted correlation across all of the studies involved in the analysis. The standard deviation of the observed correlations can then be calculated to estimate the variability in the relationship between the variables of interest.

There is one additional issue here. In many studies relying on multiple regression, there will be reported a series of models that, among other things, will reflect different independent variables and control variables. Accordingly, the *n* size of the various models may differ dramatically. It is therefore important that the sample size entered for the meta-analysis is

from the correlation matrix. Only the bivariate correlation from that matrix and its associated *n* size is relevant for the meta-analysis.

Reliability of Y/Reliability of X

As noted before, among the examples of early meta-analysis approaches are applications in industrial–organizational psychology, organizational behavior, and education. In these researches, then as now, there are often construct variables. Indeed, with the exception of demographic variables, the preponderance of work in some of these areas relies more on construct than on observed variables. Obviously, some constructs are more psychometrically sound than others. Accordingly, for *r*'s that comprise a construct variable – dependent, independent, or both – we must report the construct's reliability in the input data for meta-analysis.

By contrast, strategic management studies rarely have that character. The vast majority of that empirical work relies on observed variables. In fact, none of the extant meta-analysis approaches in strategic management studies have reported a construct variable (see Appendix B).

A question, then, arises is what is the reliability of an observed variable in strategic management studies? There are some meta-analyses that are silent on this issue, and presumably entered 1.0 (i.e., perfect reliability). We come to this conclusion because meta-analysis software normally requires some value to be entered for *Y* and *X* reliabilities. Others have clearly noted their choice, "Given our focus on objective financial variables … we only corrected for sampling error. This implicitly assumes a reliability of 1.0 and no range restriction …" (Daniel, Lohrke, Fornaciari, & Turner, 2004, p. 8; see also, Tubre, Collins, Jackson & Schuler, 2000).

We do not mean to appear critical about the choice of reliability level. Indeed the choice of 1.0 for these reliabilities, is extremely conservative. Contrary to intuitions, selecting a higher reliability does *not* result in a higher estimate of the corrected *r* in meta-analysis. All else equal, lower reliability estimates always result in higher corrected *r*'s. We will illustrate this in the "Results" section.

As reflected in Table 2, we have relied on 0.8 reliabilities for both the dependent and independent variables in the meta-analysis. The rationale is unrelated to the final estimate of *r*. Rather, it is because no variable is perfectly reliable. There is, for example, some imprecision in reporting. Return on assets (ROA), return on equity (ROE), return on investment (ROI), and return on sales (ROS), for instance, have become essentially generic, but

it is not unusual to find different estimates of them depending on the source of the data. Moreover, the closing dates for posting such estimates are not always the same; some will close on December 31, while others on June 30 in a given year. Consider another example. "Outside director proportion" is a similar generic measure, but there are several ways in which this single measure has been calculated and is apparent in the empirical literature (Daily, Johnson, & Dalton, 1999). In any case, for us a 0.8 reliability for observed variables (dependent and independent) seems reasonable and is the level on which we have consistently relied.

Range Restriction of Y and the Range Restriction of X

The issues of range restriction have become increasingly complex (for an extended discussion, see Hunter & Schmidt, 2004). At the basic level, however, the most common form of restriction – and enhancement – of range occurs by some deliberate "selectivity" in the sample. Suppose, for example, a researcher with 120 companies as a sample decides to analyze only the top 25% and the bottom 25% of those companies for some measure of financial performance. Or, suppose a researcher were interested in the relationship between some independent variables and the success of CEOs (measured by a corporate financial performance variable). But, over the period of the study, 40% of the CEOs are terminated. In this case, there is certainly a "survival" issue and the data that remain are restricted and will almost certainly affect the relationship between the independent and dependent variable for the remaining subjects.

In strategic management studies, there are always survival issues. The *Fortune* 500, the *S&P* 500, and firms trading on the various stock exchanges (e.g., NYSE, Nasdaq) are not static. Even so, range restrictions of the classical kind would be unusual. In each of the meta-analysis examples in strategic management studies for which we are aware, all range restrictions were set at 1.0 (i.e., no range restriction). In Table 2, please note that we have entered "1," no restriction in all cases for both the dependent and independent variables.

Moderator

The hypothetical study on which this meta-analysis is based examines the extent to which CEO equity holdings in publicly traded firms are associated

with an increase in firms' financial performance. We want to examine whether this relationship is informed by the "maturity" of the firm which, as noted, we operationalize as 5 or less years post-IPO or greater than 5 years post-IPO. Thus, firms' "maturity" is hypothesized to be a moderating variable.

The coding in the "Moderator" column in Table 2 represents the moderating variable. A "1" indicates that the firm is 5 years or less post-IPO; a "2" indicates firms that are more than 5 years post-IPO. Some meta-analysis software allows the researcher to note "moderators" in the way we have illustrated. When this is the case, the software may be instructed to conduct sub-analyses, i.e., in our case, a meta-analysis for all the cases coded as "1" and a separate meta-analysis for those cases coded as "2." When the software does not have this enablement, researchers can still run separate meta-analyses for subsets when necessary.

The use of the expression "moderator," instead of a "subgroup" analysis can lead to some confusion in the meta-analysis literature. For a large body of research, a moderator is ordinarily operationalized as a multiplicative variable. In a regression format, for example, one would determine if the multiplicative term provides marginal variance above that provided by the variables singly. The analog for this in meta-analysis is accomplished through establishing subgroups. Two separate meta-analyses would be conducted for the two subgroups. An estimate of the population r could be calculated for each. Then, a critical ratio test would be used to determine if the two population of r's are statistically different (we will demonstrate this in the "Results" section).

Having completed the brief descriptions of the elements of the input data for the hypothetical meta-analysis, we can address the results. While there are as many results formats as there are meta-analysis software providers, we will provide the information that would be included in virtually all of them.

META-ANALYTIC PROCEDURES AND RESULTS

The meta-analytical results in Table 3 are consistent with guidelines provided by Hunter and Schmidt (2004). Meta-analysis is the approach of choice for the hypothetical research described here to provide a synthesis of extant research that, while correcting for various statistical artifacts, allows for the aggregation of results across separate studies to obtain an estimate of the true relationship between two variables in the population.

Table 3. Meta-Analysis Results.

	Number of Correlations	Combined Sample Size	Mean True Score Correlation[a]	Standard Deviation True Score Correlation	80% Credibility Interval (lower: upper)	90% Confidence Interval (lower : upper)	% Variance Attributable to Artifacts
Entire sample	30	9,685	−0.026	0.283	−0.389:0.336	−0.112:0.059	5.74
Moderation: 5 years or less from IPO	16	2,032	0.417	0.048	0.354:0.479	0.396:0.437	80.49
Moderation:greater than 5 years from IPO	14	7,653	−0.144	0.188	−0.386:0.098	−0.061:−0.227	7.26

[a]In other meta-analysis software, this may be referred to as adjusted *r*, corrected *r*, or estimated population correlation.

Observed correlations between the variables of interest are weighted by the sample size of the study in order to calculate the mean weighted correlation across all of the studies involved in the analysis. The standard deviation of the observed correlations is then calculated to estimate their variability. Total variability across studies is comprised of the true population variation, variation due to sampling error, and variation due to other artifacts (i.e., reliability and range restriction). To control such artifacts, we relied on *Hunter–Schmidt Meta-Analysis Programs* (Schmidt & Le, 2004).

Table 3 is arranged in three rows, each reflecting a separate meta-analysis. The first row is "Entire Sample." The "Number of Correlations" column notes that the meta-analysis was conducted for all 30 correlations (consistent with Table 2, in which all 30 correlations are noted). The "Combined Sample Size" entry in Table 3 is simply the total of the samples sizes (i.e., $56 + 225 + 192 + \cdots$) as noted in Table 2.

"The Mean True Score Correlation" is the estimated mean correlation that accounts for the variance of the r's, the associated sample sizes, reliability, and range restriction. For this meta-analysis, the value is -0.026. Obviously, this is of little practical importance. More precisely, however, there is a diagnostic provided by the 90% confidence interval.[4] Notice that the interval (-0.112–0.059) includes zero. This suggests that the mean true score correlation is not significant.

The 80% credibility interval is another important diagnostic statistic. Notice its width, which is approximately 0.73 (i.e., -0.389–0.336). Koslowsky and Sagie (1993; see also, Whitener, 1990) argue that an interval of 0.11 or greater implies the existence of a moderating variable.[5] There is another indicator as well. Notice that the entry in "% Variance Attributable to Artifacts" column is 5.74%. In meta-analysis, there is an often noted "75% rule." Basically, if the variance attributable to artifacts is 75% or greater, it is unlikely – but by no means impossible – that there is a moderating variable. In this case, though, the observed value (5.74%) is not close to 75%. This joins the large credibility interval rather to strongly suggest that there may be a moderating influence on these data.

For us, however, the above indicators are satisfying because we have hypothesized the existence of a moderator. Rows two (5 years or less from IPO) and three (greater than 5 years from IPO) in Table 3 illustrate the formal tests of this hypothesis. For the 5 years or less group, Table 3 notes that there are 16 correlations (this is consistent with Table 2 in which there are 16 cases coded "1" in the moderator column) with a combined sample size of 2,032. The mean score correlation is 0.417.[6] Notice that the 90% confidence interval (0.396 – 0.437) does not include zero. Notice also that the 80% credibility

range (0.354 – 0.479) is relatively narrow and the variance attributable to artifacts is relatively high (80.49%), well above the "75% rule."

The "greater than 5 years" group (14 correlations, combined sample size of 7,653) has a different profile. The mean true score is −0.144[7] and the 90% confidence interval does not include zero. Moreover, both the 80% credibility interval (≈0.84 and the % Variance Attributable to Artifacts, 7.26%) suggest the existence of a moderating variable.

In this case, it would appear clear that the magnitudes of the *r*'s for the two groups are grossly different, 0.417 when compared with −0.144. But this is not always be the case. When the corrected *r*'s are closer, there is a test – sometimes referred to as a critical ratio – that can be used to determine if the values of these two *r*'s are different at a statistically significant level. The value that the researcher would determine with this test is essentially a *z*-score and may be interpreted as follows: a value of 1.96 would be statistically significant at the 0.05 level, 2.58 at the 0.01 level, and 3.29 at the 0.001. For these hypothetical data, the differences in the *r*'s for the two subgroups (moderators) results in a value of 11.56.[8]

The results, then, of our hypothetical meta-analysis examining the relationship between CEO equity holdings and firm financial performance can be easily summarized:

- There is no simple relationship between CEO equity holdings and firm financial performance (−0.026, ns). There is, however, some evidence of the existence of a moderating variable;
- There is evidence of a moderating effect for time since IPO. The relationship between CEO equity holdings and firm financial performance for firms 5 years of less from IPO is 0.417, is a significant relationship. The diagnostics suggest that a further moderating effect of this result is unlikely;
- The relationship between CEO equity holdings and firm financial performance for firms with more than 5 years from the IPO is −0.144, is a significant relationship. The diagnostics suggest that a further moderating effect of this result is likely.

Some Tips on Moderating Variables

The width of the credibility interval and the "75% rule" (the % Variance Attributable to Artifacts) are guidelines, not formal statistical tests. Whether there is a variable that moderates the relationship of interest is, of course, an empirical question.

Also, some caution is rightly suggested for the profligate testing for moderators. There are at least three issues, all of which are potentially related. With every separate test of moderation, there is the potential for capitalization on chance in identifying such effects. Furthermore, there is a loss of statistical power. Obviously, the number of correlations and the combined sample size will be smaller in every subgroup as compared to the overall sample. And, if the number of correlations becomes very small, the potential impact of variances between the *r*'s and sample sizes can lead to some instability in the results.

Moderators in meta-analysis are usually operationalized as a dichotomy, but they need not be. Suppose a researcher decides that there is something fundamentally different about a given relationship as a function of firms' stock exchanges, e.g., NYSE, ASE, Nasdaq. This would have no effect on the general process we have described. A suggested approach would involve running an omnibus meta-analysis (for all the studies), and then having separate tests for NYSE, ASE, and Nasdaq companies. Having said this, there will be the issues of capitalization on chance and lower statistical power, as previously described.

A number of other factors could operate as moderators. Time, for example, could be a moderator. Major interventions could also operate as moderators. In the near future, for example, corporate governance studies may be examined for relationships between pre- and post-Sarbanes-Oxley Act. For many, the passage of this Act in 2002 represents a significant "intervention" in the field of corporate governance.

OTHER ISSUES IN META-ANALYSIS AND STRATEGIC MANAGEMENT STUDIES

There are several recurring themes in meta-analysis with which researchers relying on the technique should be aware, as these constitute issues with which many reviewers will concern themselves. We develop these in the following sections.

Fixed versus Random Effects Models

Some of the meta-analysis software will require that the researcher enter a preference for whether the data be analyzed as a fixed or a random effects model. Basically, a random effects model anticipates that population

parameters may vary across studies and the variance is calculated with that assumption. Fixed model effects assume that the population parameters are invariant (Field, 2001; Hartung & Knapp, 2003; Hunter & Schmidt, 2004). Hunter and Schmidt (2004) suggest that the random effects model is the more realistic choice, as does Schulze (2004; see also, Fleiss, Levin, & Paik, 2003; Hartung & Knapp, 2003; Rosenthal & DiMatteo, 2001).

The File Drawer Problem

It has been widely discussed that meta-analysis may have an upward bias, i.e., r value estimates will be systematically higher (Begg, 1994; Duval & Tweedie, 2000; Hedges & Vevea, 1996; Hunter & Schmidt, 2004; Lipsey & Wilson, 2001; Rosenthal, 1979). This argument is based on the notion that research is selectively included in any given meta-analysis because it is almost certainly was published. In addition the general belief that empirical work presenting with statistically significant "results" is more likely to be published than its null findings counterparts and we have the foundation for the alleged bias.

While there are those who feel otherwise (e.g., Begg, 1994), we would join those who are less concerned about the file drawer issue (e.g., Hunter & Schmidt, 2004; Schulze & Wittman, 2003).[9] As noted earlier meta-analyses do not necessarily rely on data from the "main" variables of a study. More often, the raw data for a meta-analysis is derived from correlation matrixes including a host of other variables that are of little interest to the focal study (e.g., a simple correlation between two control variables). While it is true that many of these are positive, negative, and null, it is unlikely that any of them are consequential in whether the paper is accepted for publication.

That said, the problem should not be capriciously dismissed, because the impact of it will be closely associated with the number of studies included in a given meta-analysis. Consider a case wherein only four studies have been identified for a meta-analysis and we will suppose that the corrected r is 0.12. There is no doubt that the existence of, for example, two additional studies with null results that were not initially identified for inclusion in the meta-analysis could have a major impact on these results. Furthermore, if these unidentified studies also had large samples sizes – much larger than the studies that were included – then the impact will be even larger.

There is an approach, often referred to as the "fail safe" method, that may provide some comfort. The obvious question is whether the upward bias is large enough in a given meta-analysis to compromise its results. The

fail safe method developed by Rosenthal (1979; for an alternative approach, see Hunter & Schmidt, 2004) basically estimates the number of unselected studies with null reports that it would require to reduce a reported r to an unimportant level.[10] Consider the results reported in Table 3 for the moderated results, $r = 0.42$ with 16 studies. From the equation in endnote 10, we know that we would need seven additional null studies to reduce 0.42 to 0.30, 11 additional null studies to reduce 0.42 to 0.25, and 18 additional null studies to reduce 0.42 to 0.20.

But, let us return to the issue of number of studies. Suppose, instead of 16 studies with the r of 0.42, there are only four studies. Now (and we are rounding here), two null studies will reduce the 0.42 to 0.30, three such studies will reduce the 0.42 to 0.25, and four studies will reduce the 0.42 to 0.20. The lesson here, however, is less about fail safe algorithms than the exposure of relying on relatively few studies for a meta-analysis.

"Quality" of Data and Outliers

In discussions of meta-analysis, a recurring theme is the differences in the quality of studies and how, if at all, a researcher should select studies for inclusion (e.g., Cooper, 1998; Hunter & Schmidt, 2004; Lipsey & Wilson, 2001; Rothstein, McDaniel, & Borenstein, 2002; Wortman, 1994). Suggestions range from weighting schemes for studies of particularly high quality (see, e.g., Rosenthal, 1991; Schulze, 2004) to excluding poor studies altogether (Clarke & Oxman, 1999; Kraemer, Gardner, Brooks, & Yesavage, 1998). For us, the discarding of data through the elimination of studies is the proverbial slippery slope. We recognize that the point is to avoid relying on defective data. At the same time, we could imagine several researchers vetting population studies by quality and arriving at as many subsets of these studies as there were researchers and, thus, introducing an unnecessary subjective element.

Beyond that, however, there is an empirical option that does not include the discarding of data. If there are studies that, for whatever reason, are of questionable quality to the researcher, these studies may be operationalized as a "moderator." Given that, one would run the meta-analysis with all cases and then a reduced set meta-analysis excluding the troublesome studies. Whatever else might be done, researchers will have to be absolutely clear about the conventions relied on for data exclusion. We might add that from the very inception of meta-analysis Glass (1976) advocated for including all the available data. Hunter and Schmidt (2004) share this perspective. Once again, this does not suggest that the data will never be aberrant; it may

suggest, however, that such studies be coded and specifically examined for whatever moderating influence they may have on the corrected *r* of interest.

Remember, also, that meta-analysis in strategic management studies may once again have an advantage with regard to the quality of studies. As previously noted, strategic management studies commonly rely on observed variables. Moreover, the relevant data for meta-analysis approaches are typically derived from correlation matrices or converted from measures to a correlation coefficient. Accordingly, there may be relatively little discussion about the "quality" of such data.

Outliers, in many ways, have a similar character. It is notable, however, that outliers present in two ways – one as a function of the data and another as a function of data entry error. A "data" outlier is essentially another form of selection. Consider a meta-analysis in which most of the *r*'s are between 0.10 and 0.30, but with an "outlier" of 0.70. Also, assume that the *n* size for this outlier is an order of magnitude larger than the other cases. The question is whether the "outlier" should be included in the meta-analysis or not. Obviously, an outlier is not necessarily a poor study; in fact, the study may have been enviably executed. Notably, the real issue with "outliers" is not with the awkwardness in the quality of the data. Instead, the issue is almost always awkwardness in the analysis of the data.[11] Once again, we would not exclude such data. A better protocol for us is to run the meta-analysis with the full data set, and run another without the outlier(s). Frankly, we see no justification for winsorizing these data, i.e., changing the *r*, for example, so that it is only two or three standard deviations from the mean (see, e.g., Barnett & Lewis, 1994; Rousseeuw & Leroy, 2003, and Wilcox, 1997 for extended discussions of outlier data). And, even if we feel differently, we would have to consider at length what a windsorizing approach would mean even for an outlying *n* size.

A second type of "outlier," however, is entirely avoidable and that is an input error. Meta-analysis is susceptible to outliers both in the magnitude of an *r* and of a large *n* size. An input error of 0.33 instead of the correct 0.03, or a 0.33 instead of a correct −0.33 could make a large difference in the corrected *r*. Also a *n* size of 1,000 instead of the correct 100 could have similar consequences. The lesson is, therefore to double-check the data carefully.

Disclosure and Replicability

In many ways, meta-analysis is the perfect vehicle for disclosure and replicability. In principle, every study should be described in sufficient detail

such that another researcher could replicate the examination. Meta-analysis applications may be the only venue in which this could actually be done. From the onset, the actual studies and the derived data on which a meta-analysis is based is – or certainly should be – published with the article or easily available from the author(s). Accordingly, any researcher is at liberty to review every piece included in the meta-analysis for the accuracy of input entry. Moreover, those who would replicate a meta-analysis will also know the exact meta-analysis software on which the analyses were based. Finally, the meta-analysis article itself should include every assumption about, for example, reliabilities, range restrictions, data inclusion/exclusion, outliers, and the studies that comprise the subgroups for moderation analyses.

Accordingly, re-analyses with alternative assumptions is easily accomplished. Also, re-analyses including studies that were not identified or did not exist at the time of the initial meta-analysis are easily done as well. This is an exceptional advantage of meta-analysis. As noted by Lipsey and Wilson (2001, p. 168) "This process of scrutiny, and the accessibility of meta-analysis to such scrutiny, adds to its credibility and, hence, its persuasiveness in application to policy and practice." Rothstein et al. (2002, p. 538) agree, noting that "Explicitness and transparency of procedures are hallmarks of a properly conducted meta-analysis."

Independence

There are two aspects of "independence" that are potentially relevant for meta-analysis applications – independence of *r*'s and independence of samples. In the first case, the issue is what a researcher should do with meta-analysis input for a series of *r*'s that are not independent. Consider an example. Suppose, in the same study with an *n* of 200, there are four operationalizations of independent variables measuring job commitment (call them IV1, IV2, IV3, and IV4) and their relationship to a single measure of absenteeism (DV). On a correlation matrix, we will obviously see four bivariate correlations – IV1 with DV, IV2 with DV, IV3 with DV, and IV4 with DV. Because these IVs presumably capture some aspect of job commitment, we would expect that the IVs are correlated. Thus, these four correlations are not independent.

The question, then, is what do we input into the meta-analysis? One option is to input the four correlations all with an *n* of 200. Another is to combine the four studies in some way. Perhaps we could use the average *r* for the four studies; or we could use the weighted *r* of the four studies (in this

case, the weighted *r* is the mean because all four studies have an *n* of 200; obviously, this would not always be the case). And, whatever we do, what is the appropriate *n* size to input for the combined *r* – is it 200 or is it 800? Another approach is to select one of the *r*'s randomly, input that one, and ignore the others.

As one might expect, there has been extensive discussion of these issues (e.g., Cooper, 1998; Hedges & Olkin, 1985; Hunter & Schmidt, 2004; Lipsey & Wilson, 2001; Petitti, 2000). There is some, but by no means total, consensus on this issue. First, there is an irony about which any researcher should be aware. Hedges and Olkin (1985, p. 221), some 20 years ago, reminded us that any pooling of effect sizes is only sensible when the measures "*clearly* reflect the same construct" (emphasis is ours). Moreover,

> Thus pooling of correlated effects is reasonable only when the estimates are highly correlated. But this is exactly the situation in which pooling results in only a small gain in efficiency… In most cases, the gain in efficiency resulting from pooling of correlated estimates *does not justify the effort required* (Hedges & Olkin, 1985, pp. 221–222; emphasis is ours).

Curiously, then, it is in the very circumstance that when the *r*'s should be combined in some ways that the analyses will be least affected. Beyond that, Hunter and Schmidt (2004, p. 432) have noted that "while violations of the assumption of independence affect (inflate) the observed variance of effect sizes across studies, such violations have no systematic effect on the … mean *r* values in a meta-analysis."

For meta-analysis in strategic management studies, the pooling of *r*'s is a less serious issue. Few observers would argue that the financial performance indicators on which we routinely rely "clearly reflect the same construct" (Hedges & Olkin, 1985, p. 221). And, this might be argued for most strategic management studies variables. In any case, this violation – a small risk in our view – is apparently accompanied by little impact on meta-analysis results. Once again, however, to the extent to which others disagree, we would revisit the notion of disclosure and replicability of meta-analysis. We would also add that these effects, if any, are empirical questions. One can easily imagine a meta-analysis with all *r*'s entered compared to a second meta-analysis with those *r*'s in question pooled and run separately.

One last note on this issue. Suppose a researcher believes that ROA and ROI may not rise to the level of the "same construct" test, but are related. This is not necessarily a problem. Suppose, for example, that this researcher has compiled a meta-analysis database with some independent variable of interest and both ROAs and ROIs as dependent variables. There is no issue

whatsoever with using ROA and ROI as a moderator. One meta-analysis should be conducted relying on the studies with ROA and another meta-analysis conducted with the ROIs. Now, it becomes an empirical question about whether the corrected *r*'s differ on this dimension.

There is yet another "independence" question that is largely unique to meta-analysis in strategic management studies. So, often the samples on which we rely for data across a host of studies are drawn from the same population – e.g., the *Forbes* 500, *Fortune* 500/1,000, the *S&P* 500/1,000, *Inc.* 100 (now defunct). So, consider a meta-analysis designed to examine the relationship between any two variables of interest relying on 50 studies, all of which were conducted with *Fortune* 500 firms over some number of years. Is this an independence issue of consequence? Our answer is no, but there are important consequences for generalization.

One way to interpret a set of studies as we have described is as a series of constructive replications (see Lykken, 1968). Consider a group of 1,000 first year students at a given university. Suppose over the course of a year, that many researchers repeatedly use subsets of this student group for research to determine the relationship of one variable to another, call these variables "*Y*" and "*X*." Consider, further, that the various researchers use different operationalizations, which they believe capture the essence of "*Y*" and "*X*." Suppose, lastly, that there are 75 studies or so conducted over this period. Critics would aver that not all frosh classes are the same (e.g., demographic differences, regional differences, entry requirements). But, the aggregation of these studies in a meta-analysis would allow an extremely strong statement about the nature of the relationship of "*X*" to "*Y*" for this discrete group at this single university. What the described protocol amounts to is some 75 constructive replications. But, the aggregation of these studies in a meta-analysis may have very little generalizability.

The *Fortune* 1,000, for example, has a similar character. The inferential logic underlying a meta-analysis of 75 studies each drawing samples from the *Fortune* 1,000 investigating variables "*X*" and "*Y*" may be astonishingly robust. Such a meta-analysis would comprise the largest firms in the United States. Moreover, this series of studies provide a distinct methodological advantage amounting to a series of samples drawn from a discrete population, with replacement. It is true that the exact elements of the population – the largest corporations – change over time. Even so, the fundamental nature of the population is invariant. As noted, in Lykken's (1968) classic formulation, these studies amount to an extensive series of constructive replications. Repeatedly, researchers, while relying on different

operationalizations of "*X*" and "*Y*" in varied contexts, have investigated this relationship of interest.

There is, however, a downside for any sample of this nature. The results from a meta-analysis based on such data must be interpreted with some care as it would not be appropriate to generalize these findings beyond this specific sample. With the *Fortune* 1,000 example as well as the university example, we may, through the meta-analysis, be well informed about the effect sizes of the relationship of interest, but are ill equipped to generalize beyond these focused samples.

SOME LAST THOUGHTS…AND A PLEA, OR SEVERAL

It may be of some compliment to the efficacy of meta-analysis that there are 10 federally funded centers for evidence-based medicine, the charter of which is to meta-analyze the medical literature and inform clinical practice. Moreover, there are various governmental agencies, including the National Institute of Mental Health, that encourage or require a meta-analysis of the relevant research to be conducted before the agency will funds the proposed research. Hunter and Schmidt, 2004; (see also Hunt, 1997; Schulze, 2004), too, have underscored the public policy implications of meta-analysis. Surprisingly, however, there remain a few prosaic issues that do not facilitate the collection of potential meta-analysis data for this and the next generation of researchers who might seek to synthesize literature of interest to them.

Our first plea, then, is to remind everyone to include a correlation matrix in any empirical work. While there are some journals for which this is routine, there are only a few where it is universal. Moreover, in some disciplines – and no disrespect is meant – the inclusion of a correlation matrix is rare. In the major outlets for accounting and finance, for example, empirical work is rarely accompanied by a correlation matrix of the relevant variables. Consider, for instance, that we were able to identify only three meta-analysis applications in finance (e.g., Coggin, Fabozzi, & Rahman, 1993; Cooper, Ho, Hunter, & Rodgers, 1985; Fletcher, 1995). Accordingly, well executed work in accounting, finance, and economics that clearly complements research in strategic management studies (and vice versa), is difficult to include in meta-analysis applications. Our first plea: correlation matrices – each time, every time.

Our second plea: share the data. We have been involved in meta-analysis work (see Appendix B) that incorporated studies from finance, accounting, and economics. Many of these studies, and others, were published in research outlets that did not include correlation matrices. We compliment our colleagues for their kind attention to our requests to provide bivariate correlation data for their work. That said, and for many other reasons, such requests are not uniformly granted. In any case, we encourage our colleagues who might receive a request to provide data to a colleague working on a meta-analysis to please respond affirmatively to such requests. The more inclusive the input data for any meta-analysis, the better our understanding of the relationships under investigation, and its potential moderators.

Meta-analysis is not difficult, and like regression analysis, meta-analysis has proven to be relatively robust to violations of assumptions (e.g., Hunter & Schmidt, 2004; Rothstein et al., 2002; Schulze, 2004). It is probably fair to state that the greatest threat to sound execution of meta-analysis is brought about by reliance on few studies and a small *n* sizes. With this caveat, we can say that meta-analysis is extremely compliant, its assumptions are easily tested and compared, and it has demonstrated efficacy and promise in many disciplines. We hope those with little experience thus far in meta-analysis will find the hypothetical data and example provided here as a useful introduction to meta-analysis.

Should anyone elect to undertake a meta-analysis, he might begin with the guidelines summarized in Table 4. We will not review these guidelines here as they have been discussed in some detail in previous sections. We will not argue that there is a universal concurrence about these guidelines. We will, however, note that they are conservative and defensible. These guidelines do assume that there are a reasonable number of studies in the meta-analysis with reasonable sample sizes for each.

While confessing from the onset that these suggested levels are arbitrary, *n* sizes of at least 30 for each correlation seem reasonable. From a reviewer's perspective, we would think that it would be very difficult to provide a rationale for a meta-analysis with less than 10 studies. For a test of moderation, we have suggested a minimum of three studies to constitute a subgroup, but this, too, is a bit arbitrary (Dalton, Daily, Certo, & Roengpitya, 2003).

There is simply no comparison between a narrative review and a meta-analysis approach. For a means of generating a best estimate for an effect size of a body of research, meta-analysis has no equivalent. We recommend it to your attention. And, one last reminder – correlation matrix, correlation

Table 4. A Summary of General Guidelines for Meta-Analysis.

- There is no need of transforming the input values of r's.
- When it is necessary to impute the value of r, we would always set $r = 0$.
- For observed variables, rely on 0.8 for the reliability of the dependent and independent variables.
- With observed variables, it will rarely be necessary to assign a range restriction score.
- Use a conservative 90% confidence interval for the meta-analysis diagnostics.
- Use a conservative 80% credibility interval for the meta-analysis diagnostics.
- Where the meta-analysis software provides an option, rely on a "Random Effects Model".
- Assuming every effort has been made for an exhaustive search for meta-analysis input data, we would not be concerned about "file drawer" issues.
- We would neither weight nor exclude data on the basis of the quality of the study. Instead, run two meta-analyses and compare the results for the entire data set and a reduced data set without the troublesome data.
- We would not exclude outliers. Instead, run two meta-analyses and compare the results for the entire data set and a data set without the outliers.
- Only under extremely rare conditions we would have any concerns about the independence of the data and would, accordingly, be unlikely to combine data from separate rs in any manner.

matrix, correlation matrix. Alright, two last reminders. Please provide data in response to those seeking meta-analysis input correlations. Good luck on your work.

NOTES

1. Appendix A provides a compendia for the classic meta-analysis books and summary treatments.

2. We are comfortable selecting and, in retrospect, criticizing these examples because one of us was involved in both. We would note, however, that these treatments were essentially the state of the art at the time. The first meta-analysis in strategic management studies of which we are aware was published by Gooding and Wagner (1985).

3. More precisely, we know that any r between -0.17 and 17 with an n size of 120 would be just outside the standard of the 0.05 level of statistical significance.

4. While meta-analysis software will vary in the confidence intervals on which it relies, the use of 95% or 90% intervals would appear to be the standards. We have the more conservative 90% level. The 95% confidence interval in this case would have been -0.128–0.075, much wider than the 90% interval that we report, -0.112–0.059.

5. While meta-analysis software will vary in the credibility intervals on which it relies, the use of 90% or 80% intervals would appear to be the standards. We have the more conservative 80% level. The 90% credibility interval in this case would have been -0.493–0.448, much wider than the 80% interval that we report, -0.389–0.336.

6. We noted earlier that the corrected r (0.417 in this case) would be affected by the choice of reliabilities noted for the variables (see Table 1). We used the 0.8 level for the 0.417 result. The following are the results that would have been observed by other choices: 0.7 level of reliability – 0.48; 0.9 level of reliability – 0.37.

7. The equation (Fleiss, Levin, & Paik, 2003; Quinones, Ford, & Teachout, 1995) for the critical ratio test is (the corrected r for the first group−the corrected r for the second group)/sqrt of (standard error of group 1)^2−(the standard error or group 2)^2. The standard error is equal to SD/sqr of the number of correlations.

8. Even if there were a "file drawer" problem, it would be manifest for narrative reviews as well.

9. The operative equation is: the number of file drawer studies = [(number of studies in the meta-analysis)*(the corrected r/the lower r)−1]. Consider an example. Suppose there are 100 studies in the meta-analysis and the corrected r is 0.50. How many file drawer studies at r of 0 would it take to reduce the 0.50 to 0.20? So, [(0.50/0.20)−1] = 1.5. And, 1.5 times the number of studies (100) = 150. Accordingly, one would need to identify 150 studies at $r = 0$ to reduce the 0.50 from the meta-analysis to 0.20.

10. Consider, for example, a study relying on the entire *Fortune* 500 as the data. Arguably, this is not a sample, but a population. Suppose there are several "outliers" (however defined) in these data. On the basis of the quality of the data, there would seem to be absolutely no justification for excluding these outliers. They are from a population and accurate; they are what they are. On the basis of the analyses, however, there may be compelling reasons why these "outlier" data may lead to an aberrant result.

REFERENCES

Barnett, V., & Lewis, T. (1994). *Outliers in statistical data*. New York: Wiley.

Begg, C. B. (1994). Publication bias. In: H. Cooper & L. V. Hedges (Eds), *The handbook of research synthesis* (pp. 399–409). New York: Russell Sage Foundation.

Clarke, M., & Oxman, A. D. (1999). *Cochrane reviewers' handbook 4.0*. Oxford: Cochrane Collaboration.

Cochran, W. G. (1937). Problems arising in the analysis of a series of similar experiments. *Journal of the Royal Statistical Association*, *4*, 102–118.

Cochran, W. G. (1943). The comparison of different scales of measurement for experimental results. *Annals of Mathematical Statistics*, *14*, 205–216.

Coggin, T. D., Fabozzi, R. J., & Rahman, S. (1993). The investment performance of U.S. equity pension fund managers: An empirical investigation. *Journal of Finance*, *48*, 1039–1056.

Cohn, L. D., & Becker, B. J. (2003). How meta-analysis increases statistical power. *Psychological Methods*, *8*, 243–253.

Cooper, H. (1998). *Synthesizing research* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Cooper, W. W., Ho, J. L. Y., Hunter, J. E., & Rodgers, R. C. (1985). The impact of the foreign corrupt practices act on internal control practices. *Journal of Accounting, Auditing, and Finance*, *9*, 22–39.

Daily, C. M., Johnson, J. L., & Dalton, D. R. (1999). On the measurement of board composition: If you have seen one, you certainly have not seen them all. *Decision Sciences*, *30*, 83–106.

Dalton, D. R., Todor, W. D., Fielding, G. J., Spendolini, M. J., & Porter, L. W. (1980). Organization structure and performance: A critical review. *Academy of Management Review*, *5*, 49–64.

Dalton, D. R., Daily, C. M., Certo, S. T., & Roengpitya, R. (2003). Meta-analyses of corporate financial performance and the equity of CEOs, Officers, Boards of Directors, Institutions, and Blockholders: Fusion or confusion? *Academy of Management Journal*, *46*, 13–26.

Daniel, F., Lohrke, F. T., Fornaciari, R., & Turner, A. (2004). Slack resources and firm performance: A meta-analysis. *Journal of Business Research*, *57*, 565–574.

Duval, S., & Tweedie, R. (2000). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, *95*, 89–98.

Eysenck, H. H. (1952). The effects of psychotherapy: An evaluation. *Journal of Consulting Psychology*, *16*, 319–324.

Eysenck, H. J. (1978). An exercise in mega-silliness. *American Psychologist*, *33*, 517.

Field, A. P. (2001). Meta-analysis of correlation coefficients: A Monte Carol comparison of fixed- and random-effects methods. *Psychological Methods*, *6*, 161–180.

Fisher, R. A. (1932). *Statistical methods for research workers*. London: Oliver and Boyd.

Fleiss, J. L., Levin, C., & Myunghee, C. P. (2003). *Statistical methods for rates and proportions* (3rd ed.). Hoboken, NJ: Wiley-Interscience.

Fletcher, J. (1995). An examination of the selectivity and market timing performance of UK unit trusts. *Journal of Business Finance & Accounting*, *22*, 143–156.

Glass, G. V. (1976). Primary, secondary and meta-analysis research. *Educational Researcher*, *5*, 3–8.

Glass, G. V. (1977). Integrating findings: The meta-analysis of research. *Review of Research in Education*, *5*, 351–379.

Glass, G. V., McGraw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage Publications.

Gooding, R. Z., & Wagner, J. A. (1985). A meta-analytic review of the relationship between size and performance: The productivity and efficiency of organizations and their subunits. *Administrative Science Quarterly*, *30*, 462–481.

Hartung, J., & Knapp, G. (2003). An alternative test procedure for meta-analysis. In: R. Schulze, H. Holling & D. Bohning (Eds), *Meta-analysis: New developments and applications in medical and social sciences* (pp. 53–68). Cambridge, MA: Hogrefe & Huber Publishing.

Hedges, L. B. (1982). Estimation of effect sizes from a series of experiments. *Psychological Bulletin*, *92*, 490–499.

Hedges, L. W., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.

Hedges, L. V., & Vevea, J. L. (1996). Estimating effect size under publication bias in meta-analysis. *Journal of Educational and Behavioral Statistics*, *21*, 299–332.

Hunt, M. (1997). *How science takes stock: The story of meta-analysis*. Russell Sage Foundation.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis: Cumulating research findings across studies*. Beverly Hills, CA: Sage Publications.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. New York: Academic Press.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). New York: Academic Press.

Koslowsky, M., & Sagie, A. (1993). On the efficacy of credibility intervals as indicators of moderator effects in meta-analytic research. *Journal of Organizational Behavior*, *14*, 695–699.

Kraemer, H. C., Gardner, C., Brooks, J. O., & Yesavage, J. A. (1998). Advantages of excluding underpowered studies in meta-analysis: Inclusionist versus exclusionist view-points. *Psychological Methods*, *3*, 23–31.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage Publications.

Lykken, D. T. (1968). Statistical significance in psychological research. *Psychological Bulletin*, *70*, 151–159.

Meehl, P. E. (1990). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports*, *66*, 195–244.

Pearson, K. (1904). Report on certain enteric fever inoculation statistics. *British Medical Journal*, *5*, 1242–1246.

Pearson, E. S. (1938). The probability integral transformation for testing goodness of fit and combining independent tests of significance. *Biometrika*, *30*, 134–148.

Pearson, K. (1933). On a method of determining whether a sample of size *n* supposed to have been drawn from a parent population having a known probability integral has probably been drawn at random. *Biometrika*, *25*, 379–410.

Petitti, D. B. (2000). *Meta-analysis, decision analysis, and cost-effectiveness analysis* (2nd ed.). New York: Oxford University Press.

Pigott, T. D. (1994). Methods for handling missing data in research synthesis. In: H. Cooper & L. V. Hedges (Eds), *The handbook of research synthesis* (pp. 163–176). New York: Russell Sage Foundation.

Quinones, M. A., Ford, J. K., & Teachout, M. S. (1995). The relationship between work experience and job performance: A conceptual and meta-analytic review. *Personnel Psychology*, *30*, 887–910.

Rosenberg, M. S., Adams, D. C., & Gurevitch, J. (2000). *Metawin: Statistical software for meta-analysis*. Sunderland, MA: Sinauer Associates, Inc.

Rosenthal, R. (1978). Combining results of independent studies. *Psychological Bulletin*, *85*, 185–193.

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin*, *86*, 638–641.

Rosenthal, R. (1984). *Meta-analytic procedures for social research* (2nd ed.). Newbury Park, CA: Sage Publications.

Rosenthal, R. (1991). *Meta-analytic procedures for social research* (2nd ed.). Newbury Park, CA: Sage Publications.

Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. In: S. T. Fiske, D. L. Schacter & C. Zahn-Waxler (Eds), *Annual Review of Psychology*, (Vol. 52, pp. 59–82). Palo Alto, CA: Annual Reviews, Inc.

Rosenthal, R., & Rubin, D. B. (1978). Interpersonal expectancy effects: The first 345 studies. *The Behavioral and Brain Sciences*, *3*, 377–415.

Rosenthal, R., & Rubin, D. B. (1982). Comparing effect sizes of independent studies. *Psychological Bulletin*, *92*, 500–504.

Rothstein, H., McDaniel, M. A., & Borenstein, M. (2002). Meta-analysis: A review of quantitative cumulation methods. In: F. Drasgow & N. Schmitt (Eds), *Measuring and analyzing behavior in organizations* (pp. 534–570). San Francisco, CA: Jossey-Bass.

Rousseeuw, P. J., & Leroy, A. M. (2003). *Robust regression and outlier detection*. New York: Wiley Interscience.

Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology*, *62*, 529–540.

Schmidt, F.L., & Le, H. (2004). *Software for the Hunter-Schmidt meta-analysis methods*. University of Iowa, Department of Management & Organization, Iowa City: IA. #42242.

Schulze, R. (2004). *Meta-analysis: A comparison of approaches*. Cambridge, MA: Hofrefe & Huber Publishing.

Schulze, R., & Wittman, W. W. (2003). A meta-analysis of the theory of reasoned action and the theory of planned behavior: The principle of compatibility and multidimensionality of beliefs as moderators. In: R. Schulze, H. Holling & D. Bohning (Eds), *Meta-analysis: New developments and applications in medical and social sciences* (pp. 219–245). Cambridge, MA: Hogrefe & Huber Publishing.

Shrader, C. B., Taylor, L., & Dalton, D. R. (1984). Strategic planning and organizational performance: A critical review. *Journal of Management*, *10*, 149–171.

Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, *32*, 752–760.

Smith, M. L., Glass, G. V., & Miller, T. I. (1980). *The benefits of psychotherapy*. Baltimore, MD: John Hopkins University Press.

Sohn, D. (1995). Meta-analysis as a means of discovery. *American Psychologist*, *50*, 108–110.

Thorndike, R. L. (1933). The effect of the interval between test and retest on the constancy of the IQ. *Journal of Educational Psychology*, *25*, 543–549.

Tippett, L. H. C. (1931). *The method of statistics*. London: Williams and Norgate.

Tubre, T. C., & Collins, J. M. (2000). Jackson and Schuler (1985). Revisited: A meta-analysis of the relationships between role ambiguity role, conflict, and job performance. *Journal of Management*, *26*, 155–169.

Whitener, E. M. (1990). Confusion of confidence intervals and credibility intervals in meta-analysis. *Journal of Applied Psychology*, *75*, 315–321.

Wilcox, R. R. (1997). *Introduction to robust estimation and hypothesis testing*. San Diego, CA: Academic Press.

Wortman, P. M. (1994). Judging research quality. In: H. Cooper & L. V. Hedges (Eds), *The handbook of research synthesis* (pp. 97–109). New York: Russell Sage Foundation.

Yates, F., & Cochran, W. G. (1938). The analysis of groups of experiments. *Journal of Agricultural Science*, *28*, 556–580.

APPENDIX A. BOOKS AND SUMMARY MATERIALS ON META-ANALYSIS

Borenstein, M., & Rothstein, H. (1999). *Comprehensive meta-analysis*. Englewood, NJ: Biostat.

Cook, T. D., Cooper, H., Cordrah, D. S., Hartman, H., Hedges, L. V., Light, R. J., Louis, T. A., & Mosteller, F. (1992). *Meta-analysis for explanation: A casebook*. New York: Russell Sage Foundation.

Cooper, H. (1998). *Synthesizing research* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Cooper, H. M., & Hedges, L. V. (1994). *Handbook of research synthesis*. New York: Russell Sage Foundation.

Cooper, H. M., & Lindsay, J. L. (1998). Research synthesis and meta-analysis. In: L. Bickman, & D. J. Rog (Eds), *Handbook of applied social research methods* (pp. 315–337). Thousand Oaks, CA: Sage Publications.

Durlak, J. A. (2003). Basic principles of meta-analysis. In: M. Roberts, & S. S. Ilardi (Eds), *Handbook of research methods in clinical psychology* (pp. 196–209). Malden, MA: Blackwell Publishers.

Glass, G. V., McGraw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage Publications.

Green, B. F., & Hall, J. A. (1984). Quantitative methods for literature reviews. In: M. R. Rosenzweig, & L. W. Porter (Eds), *Annual review of psychology* (Vol. 35, pp. 37–53). Palo Alto, CA: Annual Reviews, Inc.

Guzzo, R. A., Jackson, S. E., & Katzell, R. A. (1987). An analysis of meta-analysis. In: L. L. Cummings, & B. M. Staw (Eds), *Research in organizational behavior* (Vol. 9, pp. 407–422). Greenwich, CT: JAI Press.

Hedges, L. W., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.

Hedges, L. W., & Olkin, I. (in press). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Huffcutt, A. I. (2002). Research perspectives on meta-analysis. In: S. G. Rogelberg (Ed.), *Handbook of research methods in industrial and organizational psychology* (pp. 198–215). Oxford: Blackwell Publishers.

Hunt, M. (1999). *How science takes stock: The story of meta-analysis*. Russell Sage Foundation.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. New York: Academic Press.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). New York: Academic Press.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis: Cumulating research findings across studies*. Beverly Hills, CA: Sage Publications.

Johnson, B. T., & Easley, A. H. (2002). Quantitative synthesis of social psychological research. In: H. T. Reis & C. M. Judd (Eds), *Handbook of research methods in social and personality research* (pp. 496–528). Cambridge: Cambridge University Press.

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, education, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist, 48*, 1181–1209.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage Publications.

Mullen, B. (2004). *Enhanced basic meta-analysis* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Petitti, D. B. (2000). *Meta-analysis, decision analysis, and cost-effectiveness analysis* (2nd ed.). New York: Oxford University Press.

Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Newbury Park, CA: Sage Publications.

Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. In: S. T. Fiske, D. L. Schacter & C. Zahn-Waxler (Eds), *Annual review of psychology*. (Vol. 52, pp. 59–82) Palo Alto, CA: Annual Reviews, Inc.

Rothstein, H., McDaniel, M. A., & Borenstein, M. (2002). Meta-analysis: A review of quantitative cumulation methods. In: F. Drasgow & N. Schmitt (Eds), *Measuring and analyzing behavior in organizations* (pp. 534–570). San Francisco, CA: Jossey-Bass.

Sauerbrei, W., Brettner, M. (2003). Issues of traditional reviews and meta-analyses of observational studies in medical research. In: R. Schulze, H. Holling & D. Bohning (Eds.), *Meta-analysis: New developments and applications in the biomedical and social sciences* (pp., 251–258). Seattle, WA: Hogrefe & Huber.

Schulze, R. (2004). *Meta-analysis: A comparison of approaches*. Cambridge, MA: Hofrefe & Huber Publishing.

Schutze, R., Holling, H., & Bohning, D. (2003). *Meta-analysis: New developments and applications in medical and social sciences*. Cambridge, MA: Hogrefe & Huber Publishing.

Schmidt, F. L., & Hunter, J. E. (2003). History, development, evolution, and impact of validity generalization and meta-analysis methods. In: K. R. Murphy (Ed.), *Validity methods* (pp. 31–66). Hillsdale, NJ: Erlbaum.

Schmidt, F. L., & Hunter, J. E. (2003). Meta-analysis. In: J. Schinka, & W. Velicer (Eds), *Comprehensive Handbook of Psychology*, *Research Methods in Psychology* (Vol. 2, Chapter 22, pp. 533–554). New York: Wiley.

Wolf, F. M. (1985). *Meta-analysis: Quantitative methods for research synthesis*. Beverly Hills, CA: Sage Publications.

APPENDIX B. META-ANALYSES IN STRATEGIC MANAGEMENT STUDIES

Boyd, B. (1991). Strategic planning and financial performance: A meta-analytic review. *Journal of Management Studies*, *28*, 353–374.

Camison-Zornoza, C., Lapiefra-Alcami, R., Segarra-Cipres, M., & Boronat-Navarro, M. (2004). A meta-analysis of innovation and organization size. *Organization Studies*, *25*, 331–361.

Campbell-Hunt, C. (2000). What have we learned about generic strategy: A meta-analysis. *Strategic Management Journal*, *21*, 127–154.

Capon, N., Farley, J. U., & Hoenig, S. (1990). Determinants of financial performance: A meta-analysis. *Management Science*, *36*, 1143–1159.

Certo, S. T., Daily, C. M., & Dalton, D. R. (2003). Underpricing: A meta-analysis and research synthesis. *Entrepreneurship Theory & Practice*, *27*, 271–295.

Combs, J., & Ketchen, D. (2003). Why do firms use franchising as an entrepreneurial strategy?: A meta-analysis. *Journal of Management, 29*, 443–465.

Daily, C. M., Certo, S. T., & Dalton, D. R. (2005). Investment bankers and IPO pricing: Does prospectus information matter? *Journal of Business Venturing*, *20*, 93–111.

Dalton, D. R., Daily, C. M., Ellstrand, A. E., & Johnson, J. L. (1998). Board composition, leadership structure, and financial performance: Meta-analytic reviews and research agenda. *Strategic Management Journal*, *19*, 269–290.

Dalton, D. R., Daily, C. M., Johnson, J. L., & Ellstrand, A. E. (1999). Number of directors on the board and financial performance: A meta-analysis. *Academy of Management Journal*, *42*, 674–686.

Damanpour, F. (1991). Organizational innovation: A meta-analysis of effects of determinants and moderators. *Academy of Management Journal*, *34*, 555–590.

Daniel, F., Lohrke, F. T., Fornaciari, R., & Turner, A. (2004). Slack resources and firm performance: A meta-analysis. *Journal of Business Research*, *57*, 565–574.

Datta, D. K., & Narayanan, V. K. (1989). A meta-analytic review of the concentration-performance relationship: Aggregating findings in strategic management. *Journal of Management*, *15*, 469–483.

Datta, D. K., Pinches, G. E., & Narayanan, V. K. (1992). Factors influencing wealth creation from mergers and acquisitions: A meta-analysis. *Strategic Management Journal*, *13*, 67–84.

Gooding, R. Z., & Wagner, J. A. (1985). A meta-analytic review of the relationship between size and performance: The productivity and efficiency of organizations and their subunits. *Administrative Science Quarterly*, *30*, 462–481.

Ketchen, D. J., Combs, J. G., Russell, C. J., & Shook, C. (1997). Organizational configurations and performance: A meta-analysis. *Academy of Management Journal*, *40*, 223–240.

King, D. R., Covin, J. G., Daily, C. M., & Dalton, D. R. (2004). Meta-analyses of post-acquisition performance: Indications of unidentified moderators. *Strategic Management Journal, 25*, 187–200.

Miller, C. C., Flick, W. H., Wang, Y., & Huber, G. P. (1991). Understanding technology-structure relationships: Theory development and meta-analytics theory testing. *Academy of Management Journal*, *34*, 370–399.

Orlitzky, M., Schmidt, F. L., & Rynes, S. L. (2003). Corporate social responsibility and financial performance: A meta-analysis. *Organization Studies*, *24*, 404–411.

Rhoades, D. L., Rechner, P. L., & Sundaramurthy, C. (2000). Board composition and financial performance: A meta-analysis of the influence of outside directors. *Journal of Managerial Issues*, *12*, 76–91.

Tosi, H. L., Werner, S., Katz, K. W., & Dunlap, W. P. (2000). How much does performance matter? A meta-analysis of CEO pay studies. *Journal of Management, 26*, 301–339.

BALANCING THEORY AND TECHNIQUE: METHODOLOGICAL ISSUES IN STRATEGIC GROUPS RESEARCH

Mark Shanley and Margaret Peteraf

ABSTRACT

Research on strategic industry groups provides numerous examples of the tensions between theory and methodology in strategic management research. After an initial explosion of largely non-theoretical, methods-driven studies led to mounting criticisms, researchers recognized the need for more theoretical guidance concerning the nature of groups and their potential influences on firm performance. This refocusing on theory has produced different research streams, each with its own methodological concerns. This chapter reviews these developments with the objective of understanding how researchers balance theory and methods in current research.

INTRODUCTION

There is a long-standing tension between theory and method in social science research. This tension is complicated in strategic management

Research Methodology in Strategy and Management
Research Methodology in Strategy and Management, Volume 2, 65–92

ISSN: 1479-8387/doi:10.1016/S1479-8387(05)02004-7

research by the relationship between researchers and practitioners. Research on strategic industry groups provides a fruitful area for examining these tensions in detail. After an initial explosion of largely non-theoretical, methods-driven studies led to mounting criticisms, researchers recognized the need for more theoretical guidance in subsequent empirical research. This refocusing on theory, however, has produced different research streams, each with its own methodological concerns. Understanding how researchers are balancing theory and methods in more recent studies is the principal objective of this chapter.

STRATEGIC INDUSTRY GROUPS

Researchers have long suspected the presence within industries of subgroups of firms whose behaviors and results differ from those of the broader industry (Hunt, 1972; Porter, 1976, 1979). Understanding these groups was important to strategy researchers and industrial economists who sought to flesh out strategy-structure models and enhance analyses of industry conduct (Ketchen, Thomas, & Snow, 1993). It was also thought that understanding these groups could shed light on the strategies of individual firms, for which complete information was often lacking (Hatten & Hatten, 1987).

Despite-early theoretical work on mobility barriers and inter-group competition (Porter, 1976, 1979; Caves & Porter, 1977), strategic group studies have, with few exceptions, remained largely data-driven, inductive, and underdeveloped theoretically for two decades. Despite the intentions of initial researchers, studies of strategic groups have focused more on the strategic choices of group members rather than on how understanding strategic groups could help to clarify our knowledge of industry structure and conduct. Some relatively early exceptions to this include Oster (1982), Tremblay (1985), Peteraf (1993), and Cool and Dierickx (1993).

The emphasis of strategic management scholars on explaining firm performance has also worked against developing ideas of strategic groups that are tied to firm decisions and behaviors independently of outcomes and results. Differences between firms and their groups were left theoretically underdeveloped. For example, while some treatments of groups see them as possessing identities and as capable of collective action (Peteraf & Shanley, 1997; Granovetter, 1998), others view them as collections of firms that are useful primarily in providing researchers with hard to obtain information about member firms (Hatten & Hatten, 1987).

Although strategic group studies proliferated in the 1980s, systematic knowledge about groups did not, such that the domain became increasingly idiosyncratic and inconsequential to strategy scholars. Study results varied widely regarding the nature and size of group effects, as well as when such effects might occur. Interpreting the accumulated results was problematic, due to a lack of clarity regarding the nature of groups and the appropriate levels of analysis for considering their effects.

THE BARNEY–HOSKISSON CRITIQUE

In a 1990 review of 27 empirical group studies, Jay Barney and Robert Hoskisson (Barney & Hoskisson, 1990, henceforth BH) criticized the entire stream of strategic group research, calling into question whether such groups even existed. BH raise this issue in terms of a tautology, namely, that researchers assume that groups exist in industries, then employ powerful algorithms that will identify groups (even in random data), leading them to the conclusion that groups exist. These criticisms had been raised by others before BH, including McGee and Thomas (1986), Hatten and Hatten (1987), Cool and Schendel (1987), and Thomas and Venkatramen (1988), but never so powerfully. In essence, BH tie theoretical problems in group research to more fundamental conceptual issues regarding the object of study. These issues also concern epistemological problems attendant how we know whether groups that we identify are reflected in the real behavior of firms and are associated with real performance consequences.

This "existence" problem is arguably not simply about whether strategic groups exist, but also concerns the extent that we can articulate hypotheses regarding groups, test for their membership, and assess whether the groups are associated with performance, then groups are "real" in the same sense as any ideas or claims are "real." Which of the many possible groups that can be identified within an industry are more important and consequential for the industry? When researchers identify groups, they arguably should be trying to determine which groups are most influential regarding the behaviors of members (and how they are influential). Whatever strategic groups happen to be, they should involve more than shared individual characteristics. Their members should in some way recognize and interact with each other so that they comprise more than just a nominal group (Dranove, Peteraf, & Shanley, 1998).

A related question concerns whether strategic groups are associated with member performance differences relative to non-members. Does group

membership confer an advantage on firms and if so, why? This performance question is a general theoretical one, related to the overall agenda of strategic management, namely, the explanation of the determinants of firm performance. It also speaks to the existence question and which of the myriad possibilities of groupings are likely to be worthy of the attention of strategy researchers – an indication of which groups "matter." This is not just an empirical issue, but also a conceptual one of clarifying which groupings have greater theoretical claims to importance and therefore merit greater empirical consideration.

One could raise similar questions about industries and firms. Indeed, the stream of research on the contributions of industries, corporate management, and business units to firm performance (Rumelt, 1991; McGahan & Porter, 1997; Ruefli & Wiggins, 2003) raises issues about the nature of these different levels and the intervening role of management. Problems regarding the conceptual foundations of firms and industries, however, are different from what they are for strategy groups, and we need not address such fundamental philosophical issues of appearance versus reality here (Russell, 1998).

Industry definition, for example, is a continuing issue in economics and a source of conflict in antitrust cases. The reality of industries, however, is supported by law, theory, and tradition, such that the existence of industries is not questioned. Rather, it is the nature, characteristics, and boundaries of industries and their impact on member firms that are the concerns of scholars, lawyers, and consultants. Nobody is questioning that *some* notion of industry influences the behaviors and outcomes of participant firms.

Firms are also social constructions, but even more than industries, we take their existence for granted. Firms, as a type of ownership arrangement, have a basis in law, theory, and institutional tradition. The modern firm evolved over centuries and continues to evolve (Micklethwait & Woolridge, 2003). The corporate form, the staple of strategy research, only received its legitimacy in the 1890s. In contrast, trusts and cartels, which are examples of inter-firm combinations in which legal boundaries do not identify where decision authority is located, were not legitimized (Chandler, 1977). Most people today interact with firms on a regular basis, whether as employees or customers, and know of firms as important actors in social and economic life. Broader groupings come to our attention much less often.

In contrast to industries and firms, intra-industry groupings have still not been entirely legitimated as objects of study. This is surprising since inter-firm groupings of various sorts have long and controversial histories and remain important in many contemporary national economies. Moreover,

inter-firm groupings can be analyzed from a variety of perspectives, extending from strategic management and industrial economics, to power and politics, culture, and political economy. Intra-industry groups may assist in linking the macrostructure of an economy and the strategic behavior of individual firms (Granovetter, 1998). This lack of legitimacy, coupled with the underdeveloped state of theory regarding inter-firm groups, is more than enough to raise questions about the "existence" of groups. To restart research on groups after the critiques of BH and others, new theoretical approaches needed to be developed. These are discussed below.

NEW APPROACHES TO UNDERSTANDING STRATEGIC GROUPS

In response to BH and other critiques, new approaches to thinking about strategic groups sought to address the theoretical and epistemological problems that had been raised about prior research. These approaches were theoretically motivated, in contrast to the largely *ex post* empirical approaches that had dominated prior work. Some of these approaches also tried to address the linkage between the conduct of member firms of a group and their performance. Many of these approaches overlap each other and the general characterizations discussed below do not preclude more complex approaches in a study. For example, while Houthoofd and Heene (2002) focus on strategic groups as embodying shared member business strategies ("business definitions"), they also note at least six different theoretical perspectives that have developed as approaches for better motivating research on strategic groups (market power, strategic choice, strategic types, cognitive, customer, and business definition).

Ketchen et al. (1993) (KTS) present a major effort to reformulate the theoretical basis of groups in terms of inductive versus deductive groups. Deductive groups here are identified in terms of common generic strategic approaches toward the business environment in a given industry (hospitals). Bantel (1998) reports on a similarly oriented study of technology-based firms. Firms may either promote new and innovative products to gain advantage, or exploit existing products more efficiently than their competitors. Firms may also choose to exploit the broad range of products and services within an industry or within more limited niches. This leads to four generic strategies that should be identifiable within any industry: entrepreneurs, prospectors, defenders, and analyzers. The performance of deductive

groups is considered in terms of the "fit" between the generic strategies of groups and the environmental conditions in which they operate. This suggests that certain strategies will be differentially profitable within some industries and less successful in others.

Although deductive groups are both theoretically motivated and identifiable, there are still questions regarding their ontological status. Are such groups, for example, clusters of independent firms with common characteristics? If so, then they are *groups* in the sense of categories. That is to say, they represent a set of individual firms with characteristics in common. In contrast, groups could be sets of firms whose members align their behaviors with one another (*or are inter-dependent*). Such a group is arguably more of a "real" entity, since it is more than the sum of its parts (Dranove et al., 1998).

As an example of how deductive groups can be more than a simple aggregation of individual firms, consider the following. In an industry with limited demand, a group of low-cost firms could recognize their mutual interest in expanding capacity ahead of the growth in demand in order to limit the entry of other low-cost potential entrants. If they choose the amount of capacity by which to expand by, say, "following the leader," then the end result would be shaped by their interactions and by their recognition of their status as a group. It would be different from the outcome that would have resulted from a set of entirely independent decisions.

The general point is that there needs to be a collective basis for action in order for a group to have a more meaningful ontological status than a category. This can often be a difficult issue to sort out empirically. For example, different managers may define the same environment differently, leading to different strategies for firms in similar "objective" situations. This may result in identifiable deductive groups, but it is unclear whether analysis of such groups adds anything over and above what could be learned from a firm level analysis. If the groups are not more than categories, it is possible that important information could be lost or distorted by relying on group-level analysis rather than firm-level analysis (Dranove et al., 1998). The real question to be sorted out is whether the managers in such groups are making their strategic choices in terms of the actions of other group members or independently. One way to address this problem is to incorporate managers (and management teams) into a study's model. This is the approach used by Pegels, Song, and Yang (2000), who look at similarities in the profiles of management teams as one of the bases on which strategic groups are formed.

Addressing the ontological and epistemological problems of strategic groups requires recognizing the need for relational behaviors among group

members (even at the minimal level of "other-regarding" behavior), if the group is to be consequential for members. Meaningful groups need to be linked somehow to the behaviors of members toward one another, such as through the status accorded by group membership. Groups characterized this way cannot just be subjective artifacts of observers or of grouping algorithms, but will instead reflect real member behaviors. As a consequence, they will have greater potential to influence member performance.

One way of developing this behavioral basis for a group is through mutual recognition and coordination. Group members recognize other members and adjust their behaviors in light of their expectations of others. Such a "cognitive" view of strategic groups emphases the judgments that need to be made by members before any sustained group interactions can occur (Reger & Huff, 1993). This is akin to a group developing an identity among members that influences a firm's decision-making and affects subsequent behaviors (Peteraf & Shanley, 1997).

It is not only mutual awareness and behavioral linkages among group members that are important. It is also important whether and how the group contributes to firm performance. An implication of the BH critique is that among strategic groups, identified, those with a systematic association between membership and firm performance will be most important.

Wiggins and Ruefli (1995) develop the idea of *performance groups* to address this issue. These are industry groupings characterized by within-group performance homogeneity and between-group performance heterogeneity. Their intuition is that the intent of strategic group theory development is to link strategic and performance groups. They propose and test whether the existence of stable performance relationships for groups provides predictive validity for those groups.

Although reminiscent of Porter's (1979) view of strategic groups, this approach has its limitations. In particular, it fails to develop the theoretical mechanisms linking groups and performance. It also fails to clarify the nature of the temporal dimension of the performance relationship. The question remains as to how stable a group must be and over what time period. How will the performance relationship change over time? How might this relationship vary across groups or across industries?

Linking group identification to member performance addresses the existence problem pragmatically, by arguing that strategic groups arise and survive to the extent to that they aid member performance. For researchers, this suggests that groups "exist" if they contribute to the explanation of firm performance, controlling for firm and industry effects (Dranove et al., 1998).

Whatever their basis in firm behaviors, it is doubtful that groups that fail to influence performance will continue to be important to managers (or receive researcher attention). Once groups are linked to performance, then the question of how group interactions contribute to performance becomes relevant. The linkage with performance also puts the idea of mobility barriers in perspective. Some idea of mobility barriers may be needed to permit the persistence of group performance effects.

This discussion has focused on the minimal bases needed to develop strategic group theory in response to the criticisms of BH and others. Numerous methodological issues are associated with these new areas of theoretical focus. The remainder of this chapter focuses on a discussion of the methodological issues raised by new approaches to strategic group theory. Although initial studies of groups were driven by issues of technique, due to a lack of theoretical clarification, subsequent methodological developments are more likely to be driven by theory, which can only be a positive development. Moreover, as we discuss in a concluding section, new research from other areas of strategy has relevance for strategic group research. In particular, studies of Asian business groups, vertical firm groupings, and strategic networks may all prove useful in future studies of strategic industry groups.

METHODOLOGICAL ISSUES RAISED BY NEW APPROACHES TO GROUPING

What are the best types of data to gather for the purpose of identifying strategic groups and what is the best starting point for such an exercise? The objective, of course, is not simply to identify groups of firms but to identify groups that change individual firm behaviors or outcomes from what they would be in the absence of groups. One approach is to begin by collecting detailed industry data and then identifying groups through clustering or other grouping algorithms. This has been the norm in prior studies, in which sets of firm-specific variables are used to group firms and establish clusters. The critical next step is to test these groups for associations with firm performance, controlling for firm- and industry-level performance effects (Dranove et al., 1998). An alternative is to test for evidence of interactions, such as collusive behavior, or changes in the degree of rivalry among group members. See, for example, the studies of how strategic groups affect rivalry by Peteraf (1993) and Cool and Dierickx (1993).

Another approach is to begin with an initial set of groups that are clearly recognized by industry members, such as distinct industry niches, high-status firms, or geographic subgroups. This is a reasonable starting point since groups, under most circumstances, need to be recognized by their members before members can become interdependent. Again, the crucial test of whether such groups are more than just categories in the minds of managers is to find evidence that they influence firm orientations, actions and outcomes from what they would otherwise be. Data for such groups may come from institutionalized sources or database categories, but it may also be necessary to rely on informed participant reports, journalistic accounts, court documents, and other sources for information on member interactions, the details of which are unlikely to reach regular industry databases.

Although a "real" group, with strong influence on firm behaviors and outcomes, would likely be apparent to members, industry groups that are apparent to members may still not be "real" in our sense of this term. Validation of observed group-level effects may well require a balancing of qualitative and quantitative approaches. A validation approach that recognizes the importance of statistical power is certainly desirable (Ferguson & Ketchen, 1999), although the data demands of such an approach are likely daunting in terms of the number of firms and groups needed, especially if split-sample approaches are used (Punj & Stewart, 1983).

It is tempting to imagine that groups identified from archival sources will converge with those that are well recognized by industry practitioners if the groups are "real." Nath and Gruca (1997) report such a finding in their study. One should be careful of such expectations, however, for several reasons. To start with, the type of groups that will be identified will likely be influenced by the researcher's assumptions about what constitutes a strategic group. If groups, for example, are taken to reflect shared business strategies, as suggested by Porter (1979), then the groups that are identified will reflect patterns of practice in the industry and should converge with the judgments of industry participants.

In regulated industries of the sort that Nath and Gruca (1997) studied, the tendency for convergence will be even stronger, since regulations may encourage categorization and at the same time further constrain the perspectives of industry members. None of this, however, tells us anything about whether the groups affect the firms involved in any meaningful way. They may be nothing more than sets of independent firms that happen to follow similar strategies. Moreover, if one assumes that firm strategies within a group differ (McNamara, Deephouse, & Luce, 2003), or that

member attachments to a group vary in strength (Peteraf & Shanley, 1997), then there is little reason to expect that participant-identified groups will converge with those identified through archival materials.

WHAT VARIABLES AND MEASURES SHOULD BE USED TO ESTABLISH GROUPS?

Identifying strategic groups involves more than just choosing initial groups and data sources. Depending on how one theorizes about groups, the measures used to group will have much influence on the groups that are identified. There has been a remarkable variety of variables used in strategic group studies. These have ranged across the traditional categories of business (production, finance, marketing, R&D, etc.) and have covered most aspects of business activity and levels of aggregation. KTS provide a detailed enumeration of nearly one hundred such variables from prior studies.

Having a detailed list of possible variables does not solve the problem of which to choose or how to measure them. For example, why should production capabilities be included in a grouping study? What is the theoretical mechanism by which this characteristic affects group behavior? What other dimensions of firms need to be assessed? How should they be measured? How many grouping variables are needed? Is only one sufficient if it provides evidence for true group-level interaction or outcomes?

These questions suggest that insufficient consideration has been given to which variables are theoretically appropriate for the purpose of distinguishing true from spurious groups (Dranove et al., 1998). Moreover, issues related to time and group dynamics remain unresolved. For example, should variables reflect activities that are easily changed by management or should they reflect long-term commitments of firms? How frequently must interactions among group members take place before the group is considered consequential? How long a period must groups persist as interacting bodies of members to be meaningful? Clearly, these questions cannot be answered adequately without a theory of strategic groups that is sufficiently well developed.

One way to see the need for theory in analyzing groups is to consider a study of a strategic group in which the identification of the group is not in question. Browning, Beyer, and Shetler (1995) study the government sponsored industry consortium SEMATECH as a vehicle for building

cooperation, even among direct competitors, on joint and risky innovative projects. This is a clear example of a strategic group, and one that attained considerable publicity and notoriety. The identification of group members was a matter of historical record and thus not at issue for the study.

The intent of the study was to consider how the underlying culture of the consortium developed and how the participants structured their interactions to foster innovative projects. The variables in the study thus were not related to characterizing the firms comprising SEMATECH and any shared strategies among them. Rather, they were designed to characterizing the nature and quality of interactions among the member firms, which is how the consortium added value. A grounded theory approach was employed. The core analytic categories of the study concerned culture, conflict, the nature of social ties, the extent of inclusiveness in the consortium, the nature of member contributions, beliefs in the openness of interactions, and other such factors.

Several things about this study are worth noting. Firstly, the fact that the group provided added-value to its members suggests that the group was more than a simple aggregation of firms. Without the group and the interactions among its members, this added-value could not have been attained. This evidence of significant group-level interaction effects suggests that the consortium was indeed a "true" strategic group (Dranove et al., 1998). Secondly, note that the types of variables that were collected capture the context for and nature of the interactions among the members. These variables are conspicuously absent from the long list presented in KTS, mentioned above. And yet, they are precisely the sort of variables that need to be measured to ascertain whether what looks like a group is actually functioning as a group. Is the group truly a *group* or merely a *category* in which similar types are grouped together?

A major emphasis of prior research has been on identifying patterns of similarities among group members, following from Porter's (1979) definition of strategic groups as a set of firms with shared strategies. This has been a popular approach to strategic grouping in prior studies, although it is not clear whether such an approach has value. Hatten and Hatten (1987) see groups as analytical conveniences that will assist researchers when detailed data on individual firm strategies is unavailable. Since it is often difficult to get detailed information on a firm's strategic decisions, the use of a strategic group may provide a useful proxy for the unseen strategic decisions of the firm. Although that may be correct, Dranove et al. (1998) show how using groups as analytic conveniences may actually distort rather than sharpen analysis. What is less clear is how a "shared strategies" definition of groups

imparts any additional explanatory power to researchers when firm-level data are available.

A "shared strategy" approach to groups can be motivated by the contingency argument that members adopt similar strategies as a common response to common environmental conditions or changes in those conditions. A contingency approach, however, needs more than a specification of the strategic decisions of members and the common environmental conditions they face, since managers are part of any contingency model and can exercise strategic choice. Managers (and top-management teams) can differ in how they interpret their environments and in how they act on those interpretations. This makes it necessary to explain why the managers for a set of firms chose to interpret the environment in a particular way. This means that one needs a pattern of similarities among management teams in a group as well as commonalities in shared strategies (Pegels et al., 2000).

In addition, unless the management teams of member firms made their strategic decisions with the actions of other members in mind, it is unclear what such a group adds to an explanation of firm behavior. Unless the variables relating to firm strategies and behaviors are relational, there is no way to tell whether such a group is real or nominal, since a similar grouping would result if all firms made their decisions independently of any other firm. If this is so, then any group that is identified is a group in name only.

One should also consider which types of measures might be most appropriate for grouping. A common approach has been to group firms according to the *levels* of certain variables, such as whether a firm's products were more or less differentiated, higher or lower quality, or whether a firm made large or small investments in assets (Ferguson, Deephouse, & Ferguson, 2000). Grouping on the basis of levels, however, is not inherently relational and will not guarantee that the identified groups will be influenced by actual interactions among members. Correlations and other associations among level variables will not indicate real interactions, unless the data and variables themselves have the potential to imply real relationships. A finding that measures of correlates of group membership are associated does not by itself suggest why the association is present or whether it relates to the strategic behaviors of members.

An alternative to using the levels of variables is to focus more on the variation of key variables around group means, relative to these measures for other groups. The idea is that on some dimension, group members would be more similar in their behavior to each other than they would be to members of other groups in the industry. This is related to the issue of whether group members may be expected to compete more or less intensely

with other group members (Porter, 1979; Peteraf, 1993). For example, if group members are expected to compete more aggressively with each other than with firms from other groups on price, we would expect a lower variation in prices charged within groups than across groups, as well as lower margins for firms in a competitive group. If, however, groups are collusive in nature, then reduced variation in prices might be expected, although this would come with higher rather than lower margins for group members. Higher and more stable margins may also be accompanied by greater variation in product differentiation, as group members compete on product characteristics and services rather than price (McNamara et al., 2003).

How Many Groups to Identify?

Another theoretical issue with methodological implications that has arisen with new explanations of strategic groups is the number of groups to identify. This is an issue on which clustering algorithms are silent, since one can choose the number of groups to identify. Some hierarchical clustering algorithms provide a full range of cluster solutions in their results. The choice should not be based on the details of algorithms, but on more theoretical grounds.

Should all the firms in the industry be grouped, or can there be some firms that are not part of groups? Porter (1979) suggested partitioning the industry into a series of groups. But some, indeed many, firms in the industry may not be in meaningful strategic groups and including them in a catch-all residual group would be misleading. McNamara et al. (2003), for example, note in their study of Minneapolis commercial banking, that some firms may be solitary and not part of any strategic group.

The choice of how many groups to identify has important theoretical consequences. A decision to group all firms (or as many firms as possible) within an industry involves an assumption that the industry in questions is structured by subgroups that can be identified by researchers. Without more clarification, this assumption can readily lead to the tautology raised above of assuming groups and then confirming that assumption through cluster results.

An alternative to complete grouping is to focus on the importance of a few strongly identified groups. Such a focus requires consideration of the criteria for whether groups are important, which raises questions regarding the basis for group identification in the actual behavior of members and the recognition of groups by industry participants. This sort of focus also

requires a deeper understanding of the industry context and a better theoretical basis for grouping and performance hypotheses. However, it is likely more plausible than a focus on complete grouping. Examples of industries influenced by one or a few groups are easy to find and include such well known industries as airlines, brewing, automobiles, and pharmaceuticals. Industries that can be fully partitioned into consequential subgroups are not easy to think of, although industries that mix national and local geographic markets, such as banking, newspapers, radio stations, and real estate brokers, may serve as examples.

A related issue to the number of groups is the temporal dimension of grouping. There is very little theoretic guidance in this area. Peteraf and Shanley's (1997) work on the formation of groups and their identities is an exception. There has been increasing empirical attention on the temporal dimension of groups. See, for example, the papers by Cool and Schendel (1987), Mascarenhas (1989), Wiggins and Ruefli (1995), and Nair and Filer (2003).

Despite some progress, significant questions remain unanswered. Are strategic groups assumed to be as long-lasting and stable as firms and industries? Given the lack of legal and institutional support for groups, it is hard to expect them to display the same relative stability as firms and industries. How do strategic groups form and how do they evolve? Will groups be more likely to arise at some points in an industry's history (start-up; external threats; decline) and less likely at other points (growth; maturity)? How long must groups exist to be identified and merit researcher attention? Would a short-term alliance among domestic firms to combat the entry of a foreign competitor into the industry constitute a strategic group? What about a temporary project-based alliance in the film industry? This is only a small sampling of questions regarding strategic group dynamics and temporal persistence. But the lack of answers suggests that more fully developed theory in this area would help researchers develop new research designs.

How to Identify and Evaluate Mobility Barriers?

The notion of mobility barriers has been linked to research on strategic groups from the beginning (Caves & Porter, 1977; Porter, 1979). The idea is that mobility barriers restrict entry into strategic groups in the same way that entry barriers limit entry into industries. Mobility barriers delineate the boundaries and preserve groups that are already formed. This makes them

important for insuring some stability and durability for strategic groups. Mobility barriers do not, however, cause groups to form or create a linkage with performance. This makes them necessary but not sufficient conditions for strategic industry groups, especially those associated with performance differentials. Although there is little debate over the nature and function of mobility barriers, how they should be identified and measured is far from clear.

Mobility barriers often involve costs of entry that effectively give group members a cost advantage over potential entrants. Accordingly, an obvious approach for investigating mobility barriers is to identify entry costs and seek an association between such costs and the ease of entry into a group. Unfortunately, the problems with this approach are numerous. First, relying on observable costs may be misleading. The effective cost of entry may differ from what public accounting data suggest. Moreover, the barrier may be due to intangibles, such as reputation or image, which are hard to measure. Or they may be due to economies of scope, which present other types of measurement problems.

A second problem is that even if barriers can be estimated, it may be difficult to link them to the profitability of group members. This has been a problem with industry studies of concentration or scale economies (Schmalensee, 1989) and there is little reason to expect any greater success with strategic groups. The methodological issue is how to sort out the potentially conflicting reasons behind an identified linkage between group membership and profitability.

How should mobility barriers be treated for industries with multiple groups? Is it reasonable to assume that all the groups in an industry will have similar types of mobility barriers? Although there may be such industries, it is also quite possible to think of industries where groups might have different mobility barriers. One group, for example, may benefit from scale and scope economies. Another may benefit from focusing on relationships with customers in critical segments. Still others may have beneficial geographic locations. The analysis is greatly complicated if a diverse set of mobility barriers is expected in an industry. This complexity is perhaps why Mascarenhas (1989) stressed the importance of industry context in understanding strategic group formation.

Another issue is what constitutes a *group* mobility barrier as opposed to an *individual* mobility barrier, known as an *isolating mechanism* (Rumelt, 1984). If all the firms in a group have made significant sunk investments in production capacity, does that constitute a mobility barrier for that group? It might be that the firms in the group made their investment decisions in

light of other group members' choices. Alternatively, it might be possible that if the firms were few in number they could explicitly or tacitly collude behind the barrier. Hayes, Spence, and Marks (1983) provide an example in their extensive case study of the investment banking industry. In this situation, the shared investments were accompanied by other conditions that increased the costs of entry into the group and facilitated collusion among group members.

In considering the role of shared investments as mobility barriers, it is worthwhile to consider why a group would be important if all such investments were made individually and without any reference to the activities or capabilities of other firms. If this were so, the researcher would not need to know anything about the group – analyzing individual firm investments across the industry would lead to the same result. For a common profile of investments to suggest a group, there needs to be more of a collective basis for such investments. Otherwise, the group is inconsequential *at the time that the investments are made*. On the other hand, if the barrier formed by common investments leads to subsequent group-oriented activity, it could be the basis for group formation.

This suggests that even with respect to mobility barriers, it is the interactions among group members that matter. These may result from the unintentional erection of a mobility barrier through, say, similar investment choices, but take place after the fact. Or they could come about in the process of erecting the barrier itself. As an example of this, consider investments in entry barriers made jointly by partnering among group members.

One approach that has been used to research mobility barriers is to assume their existence and look for a sustained linkage between group membership and profitability for validation. Wiggins and Ruefli (1995) do just this, noting that if there were not mobility barriers at work, then it would not be possible for group profitability to persist, since entry would level off excess profits. Put another way, in the absence of persistent performance, whatever mobility barriers are at work in the industry are ineffective.

The problem with this is that it confounds group effects with individual firm effects. If the barriers are actually isolating mechanisms (Rumelt, 1984) that are common to a set of independent firms, rather than mobility barriers, the results will look the same. Assuming the presence or absence of mobility barriers will certainly simplify a researcher's tasks, since the identification and validation of specific barriers will be unnecessary. This simplification, however, comes at the cost of confusing group identity with

the benefits and costs of group membership. Focusing on performance groups also limits our ability to understand the interactions among group members that supposedly generate performance effects. In the extreme, if the only groups that exist are performance groups, then why is it necessary to even consider shared strategies, mobility barriers, competitive interactions, and related topics?

How to Identify and Evaluate Competitive Interactions?

A common thread in new approaches to strategic groups is that there must be some group-level effects that are distinct from the effects of membership in the broader industry and from firm-level effects (Dranove et al., 1998). For a group to be meaningful, membership needs to be associated with some collective interactions. This is opposed to considering groups as analytical conveniences that can reveal information about member firms in the absence of firm-level information (Hatten & Hatten, 1987).

How should such interactions be identified, assessed, and linked to a group? One could start by considering behavioral linkages among members through, say, communications or different forms of transactions. Such linkages could even provide the basis for group identification, if they were used as the data in grouping algorithms, as is common in network research (Gulati, 1995). Groups would be indicated by dense patterns of relationships, relative to the rest of the industry, and behaviors among group members would display dynamic regularities not observable among other firms (Nair & Filer, 2003; Human & Provan, 1997).

Even the use of relational data, however, may not be sufficient for identifying the boundaries of groups and pinpointing their effects. Grouping algorithms can produce different results, depending on the data used and criteria chosen for defining groups. Moreover, patterns of interactions among a set of firms do not necessarily imply behaviors different from what would be expected on a firm-level basis alone. For group effects to occur, there also needs to be some recognition by members of the group and its identity and some orientation of individual firm behaviors in terms of the group (Reger & Huff, 1993; Peteraf & Shanley, 1997). Ferguson et al. (2000) consider this basis for groups in a different direction by emphasizing how reputation and status structures can make the group real in the view of industry observers, a situation which can bring benefits to members even in the absence of more significant interactions among members. This also represents a legitimate interpretation of group-level effects. The effects

extend beyond a simple summation of individual firm-level effects. They are arguably due, at least in part, to the existence of a group.

Traditional approaches to the study of market power effects and collusion seem to combine both behavioral linkages and collective orientation (Dranove et al., 1998). For example, explicit collusion necessarily indicates both the recognition of a group by its members and collective action in setting prices and production. Spar's (1994) study of commodity cartels takes this approach, focusing on the economics, politics, and organization of the cartel, in conjunction with how its actions are coordinated by its leaders. In all of her cases, including those cartels that were unsuccessful, clear actions were taken to promote collective behavior – actions that were understood by the firms involved and oriented toward influencing industry outcomes.

Explicit attempts at cartelization are not the only plausible collusive bases for competitive interactions within an industry group. There are a variety of actions associated with tacit collusion that are similarly relevant, including pricing rules of thumb, price posting, and price leadership by a dominant firm (Dranove et al., 1998; Scherer & Ross, 1990). There could also be coordinated collaborative behaviors aimed at improved efficiency or innovation, without the focus of traditional collusion on managing prices and production (Browning et al., 1995; Bresser & Harl, 1986). The overall point of studies focusing on the conduct of group members is not to suggest that there is only one variety of competitive interactions that needs consideration. Rather, it is that without competitive interactions of some sort, whether competitive, cooperative, or even collusive, it is difficult to see how group membership has any possibility of influencing firm behavior or performance. Ferguson et al. (2000) study provides the exception to this rule by looking at the effects from *external* recognition of a group.

A methodological conclusion from looking at conduct among group members is that such a focus may drive research toward industry-specific studies. Cool and Dierickx's (1993) study of rivalry in the pharmaceutical industry and Peteraf's (1993) study of rivalry in airline markets provide illustrative examples. Researchers may also find it beneficial to employ a combination of qualitative and quantitative research designs. In many cases, the data requirements for looking at the significant interactions among all the firms in an industry will be considerable. Moreover, making sense of the patterns that are identified through such analysis will also require considerable industry and firm-specific knowledge. Spar's (1994) analysis of international cartels in four commodity industries – diamonds, uranium, gold, and silver – illustrates the richness of data needed even for relative

simple industries. Her study also demonstrates the formidable tasks involved in attempting to analyze the situations in different industries beyond the level of clinical judgments and comparisons. Scott Morton (1997) provides 50 years of historical data in her study of the conditions in which British shipping cartels acted like cartels and initiated price wars in response to new entry.

Among the few studies that have considered competitive interactions, most consider interactions as a defining aspect of groups, but consider the linkage between interactions and performance separately. For example, Jobber and Lucas (2000) have performed a qualitative analysis to clarify the links between environmental conditions and firm actions in a strategic group within the Canadian computer industry (without analyzing links to performance). Similarly, Nair and Filer (2003) have looked at the interactions of firm behaviors within groups in the Japanese steel industry, also leaving the question of the association with performance. They found mixed results about how group members respond to exogenous shocks. Among firms whose strategies were significantly oriented toward their group (or *cointegrated*), some firm responses to shocks converged among group members, while the responses of other firms diverged from group behaviors. The results are strongly consistent with the predictions of Dranove et al. (1998) about the interaction of firm and group influences in a situation where strategic groups are present.

HOW TO MEASURE GROUP CONTRIBUTIONS TO FIRM PERFORMANCE?

A tension in thinking about strategic groups and performance concerns sorting out the endogeneity in performance issues to clarify whether performance differentials are characteristics or consequences of a group. On the one hand is the idea that the group influences performance such that membership helps the firm to perform better or sustain its profitability longer. On the other hand, group members bring industry backgrounds and performance histories with them when they begin to associate with one another, raising questions regarding the direction of any causality. The simple identification of an association between groups and firm performance might not imply anything about the group's influence on what members actually do.

Perhaps the best example of an empirical paper that attempts to sort out these issues is that by Nair and Kotha (2001). Following the methodological

suggestions of Dranove et al. (1998), they have tested the proposition that group effects might be spurious, due to unidentified firm effects. By choosing an industry with well-defined groups and applying econometric techniques to control for both environmental and firm-specific effects, they provide convincing evidence that group membership in the Japanese steel industry, over a well-defined period, was significantly associated with firm performance.

Studies such as this legitimate research on strategic groups and encourage further efforts. Other approaches could also be used to search for group performance effects. One approach might be to compare firm performance in groups against firm performance in alternative grouping schemes or in the absence of grouping. Since it is impossible to observe these alternative arrangements simultaneously with the observed group, testing group performance contributions must be done indirectly. For example, a longitudinal analysis of performance over a period in which group influence and environmental conditions both vary might allow one to observe the contribution of groups to performance. Another possibility is to model alternative groups of firms in the industry through Monte Carlo or bootstrap sampling to generate alternative arrangements against which observed performance can be tested (Dranove & Shanley, 1995).

Along with study design issues are concerns about what variables should be used in assessing performance. Should performance be evaluated in terms of over-all profitability or on more specific dimensions? Should intermediate outcomes be clarified, such that group interactions contribute to performance through them and not directly? Prior studies provide little guidance on these questions. KTS, for example, provide a list of nearly 50 performance variables from six principal constructs: sales, equity and investment, assets, margins and profits, market share, and overall performance, and employ measures from each construct to maintain consistency with prior research. Without some guidance for selecting performance variables, however, it will be difficult to interpret the linkages identified between group membership and performance.

Providing such guidance requires the development of better theories about how strategic groups influence firm performance. Consider a simple example. Suppose a group of firms is composed of firms possessing the same sort of critical resources. Having resources in common would not, necessarily, make this group more than an aggregation of individual firms. Now suppose that these firms attempted to collude regarding the price of the critical resource. This would certainly involve collective interactions and make this set of firms a true strategic group. In this example, how should the

performance contribution of the group be assessed? A starting point would be to determine how successful the collusion was. Specifically, the intermediary variable of the price of the critical resource could be examined in terms of its level and variation. Whether or not the group was effective at managing the price could be evaluated.

The effectiveness of collusion in the group, however, is a separate issue from the profitability of the group. The group could be very effective at colluding on price, but the costs of doing so, in terms of decision time, enforcement activities, fines from regulators, or just inefficient use of resources, could well exceed the benefits of collusion. Therefore, the group might not contribute to overall profitability (or may even detract from it), even though it was successful on the intermediate variable of price stability.

It is important to choose variables to match one's expectations regarding interactions. In the example above, if the anticipated logic of a group's interaction involved collusion, then variables relating to prices and margins would be important. Collusion has less to do with operational efficiency, innovation, or new product development. These would be relevant, however, if more cooperative interactions were anticipated (Bresser & Harl, 1986). For example, in a study of inter-corporate technology alliances and their influence on performance, Stuart (2000) looks at innovation rates and sales growth as intermediate and long-term performance variables.

The temporal dimension of strategic group performance is also important but theoretically underdeveloped (Zaheer, Albert, & Zaheer, 1999). When and under what conditions will group membership be associated with performance? When will firms find collective action preferable to independent action? Collective action might most likely be found in new industries struggling for customer acceptance and legitimacy, or in industries with weak competition and appropriability conditions. This is because individual firms in such industries will be especially vulnerable to competitive pressures, due to their small size, lack of resources, management uncertainty, and other factors (Sakakibara, 2002). As industries and firms develop, however, collective action may be less attractive relative to the benefits of independent action.

Shocks to the industry may also stimulate groups, although it is not clear that collective action under such conditions will be sustainable (Nair & Filer, 2003). A long-term industry decline may be associated with more strategic group activity to maintain current levels of business, or to lobby regulators for protection (Marsh, 1998). Similarly, threats to the legitimacy

of an industry may stimulate greater group-oriented activity as firms seek succor from group association (Peteraf & Shanley, 1997).

Relatively, few studies have attempted to link interactions within and between groups with performance. Cool and Dierickx (1993) provide an example with their analysis of U.S. pharmaceutical firms. They found that profitability is negatively associated with rivalry, whether within or between groups, as measured by Herfindahl-like indices that served as proxies for competitive interactions. They also found that relationships with performance change, as patterns of rivalry change, and that the linkage between group conduct and performance could be positive or negative. This is consistent with the theoretical expectations of Peteraf and Shanley (1997).

Peteraf (1993) examines rivalry within and across groups, using data from the airline industry. She observes that the pricing behavior and price/cost margins of airline monopolists vary according to whether the potential entrants that they face come from the same strategic group as the monopolist or a different one. The evidence indicates that rivalry is greater across groups than within groups, as predicted by Porter (1979).

Stuart (2000) finds a relationship between the characteristics of partners in technology alliances and a firm's performance. He argued that this may result from improvements in status or reputation stemming from an alliance. This minimal effect will accrue to members of a relationship, even if the intended dynamics of the relationship did not occur, which happens frequently in technology alliances. Of course, if the intended objectives of a relationship were realized, the performance effect would be stronger. This is consistent with differences in strength between strategic groups, with performance effects likely when group effects are stronger (Peteraf & Shanley, 1997).

McNamara et al. (2003) offer an ambitious design for linking group interactions and performance. They study firm differences within groups in a local commercial banking market. They find core and secondary firms within groups. Their results suggest that secondary firms outperform the core firms in their groups, as well as firms not associated with particular groups. This suggests that secondary firms may be most able to balance firm-level and group-level demands in a competitive setting.

Short, Ketchen, Palmer, and Hult (2004) report on a multi-level study of firm performance that builds on the "contributions to performance" research cited above and is consistent with the framework in Dranove et al. (1998). They use hierarchical linear modeling to simultaneously estimate firm, strategic group, and industry level influences on short- and long-term

measures of performance. Using longitudinal data for a sample of 1,163 firms in 12 industries, they assess the levels' explanatory power. They find that industry-, group-, and firm-level analyses all help explain performance, and do so in different ways, indicating that the strategic group level is important.

RELATED RESEARCH

Along with ongoing research on strategic groups, there has also been considerable work done in related areas whose insights could be productively applied to strategic groups. These areas include strategic networks, strategic alliances and joint ventures, geographical clusters, and vertical business groups.

Research on strategic networks has been addressing the broad question of how the context of firm actions, in terms of the place of the firm within a large network of organizations, influences firm action. The typical methodological approach of work in this area has been to start from dyadic relations between firms and build a network of overlapping and linked dyadic relations. This is directly relevant to strategic groups research, since groups, in network terms, are sets of firms with relatively dense interaction patterns relative to non-members. Network techniques could prove very useful for understanding competitive interactions within and between groups.

The network literature is extensive and cannot be reviewed here in any detail. Osborn and Hagedoorn (1997) provide a very informative overview of strategic networks research. Two studies that are especially relevant for strategic groups and their dynamics are those by Madhavan, Koka, and Prescott (1998), on how industry changes affect intra-industry networks, and the study by Rowley, Baum, Shipilov, Greeve, and Rao (2004) on clique structures within industry networks, which are strikingly similar to strategic groups.

Research on strategic alliances and joint ventures has developed in a similar fashion, but with a specific focus on particular and more formal modes of inter-firm governance that fall short of ownership. The literature, like that for networks, is large, but two examples show the potential for linkage with strategic group studies. Gulati's (1995) study of the multiple bases for choice in inter-firm alliances provides an excellent review of the issues, in terms that are very consistent with strategic group research. Garcia-Point and Nohria (2002) study the dynamics of alliance formation in

the global automobile industry. They conclude that the appropriate level of analysis for studying grouping and imitative processes is at the level of strategic groups or niches within an industry rather than at the industry level.

A final category of related research includes studies of Asian business groups, geographical clusters, and vertical groups, all involving buyer–supplier relations. Strategic groups have been typically defined in terms of horizontal industry groups. Given the increasing importance of differentiation-based competition, along with the fact that many business units are embedded within larger corporate actors, it seems reasonable to consider vertical linkages as a potential basis for grouping. Shanley and Peteraf (2004) provide a framework for looking at vertical groups and a review of related research.

Perhaps the most visible form of vertical group is the Asian business group, which is common in various forms in Japan, Korea, China, Indonesia, and India. Granovetter (1998) notes the limited research on these large diversified groups with some amazement, since they control economic transactions in a large part of the world economy. Studies of these groups are increasing, however, including attempts to estimate their performance effects on member firms. Khanna and Rivkin (2001) provide a good review that addresses issues directly relevant to strategic groups, including the nature of these business groups, what contributes to their cohesion, the nature of within-group interactions, and the possible bases for group performance effects.

CONCLUSIONS

In reviewing recent research on strategic groups, we see reason to be encouraged. Scholars are asking reasonable theoretical questions about such groups and then crafting ambitious study designs in attempting to shed light on them. New theoretical directions have been clearly identified, work is proceeding, and progress is being made. In addition, large research streams in other areas related to strategic groups, such as intra-industry networks, strategic alliances, and vertical business groups, are also developing swiftly. Progress in these areas holds out the promise that research on strategic groups can be enriched by this progress as well. Based on the discussion in this chapter, we have summarized some suggestions for researchers who wish to make further progress in their strategic group research. These are presented on Table 1.

Table 1. Suggestions for Researchers to Improve Strategic Groups Studies.

1. What is the best starting point?
   - Start with a set of groups that are clearly recognized by industry members, such as industry niches, high-status firms, or geographic subgroups.
2. What variables and measures should be used?
   - Select variables linked to the theoretical motivation for grouping.
   - Select variables and measures appropriate to the behavioral basis for grouping, e.g., characteristics, relationships, and changes.
3. How many groups should be identified?
   - Focus on a few strongly identified groups. Additional groups are likely to be inconsequential.
4. How to identify and evaluate mobility barriers?
   - Identify potential barriers independently of group profitability.
   - Expect different groups in an industry to have different mobility barriers.
5. How to identify and evaluate competitive interactions?
   - Consider the nature of the interactions and look for appropriate measures.
   - Look for evidence of a collective orientation.
   - Focus on industries with longitudinal data, where possible.
6. How to measure group contributions to firm performance?
   - Have specific rather than global performance dimensions in mind.
   - Assume longitudinal variation in performance relationships.
   - Use rich data contexts to model performance relationships; exclude alternative hypotheses, and capture longitudinal patterns.
   - Make use of periods of discontinuity to assess performance contributions.

Admittedly, the pace of progress in the area of strategic groups has been slow. But it is wise to remember that for a long time strategic group research was one of the most active but least productive in strategic management. It took considerable effort to reflect on the limitations of this research and think through new directions in which group studies could advance. This reflection has taken time. Given the progress to date, it is likely that patience and continued efforts by researchers will yield a high return.

One of the reasons why strategic groups research has been slow to develop is that it is a relatively difficult area to study, featuring as it does groups without the same status as industries and firms, operating in varying temporal dimensions, and embodying varied and complex logics. Linking group dynamics to performance further complicates a difficult set of research issues. The theoretical directions for studying these issues have, in recent years, become clearer. This suggests that, while there are still formidable difficulties involved in studying strategic groups, the research can proceed on a far sounder basis. We see this as a basis for optimism.

REFERENCES

Bantel, K. A. (1998). Technology-based, "adolescent" firm configurations: Strategy, identification, context, and performance. *Journal of Business Venturing*, *13*, 205–230.

Barney, J., & Hoskisson, R. (1990). Strategic groups: Untested assertions and research proposals. *Management and Decision Economics*, *11*, 187–198.

Bresser, R., & Harl, J. (1986). Collective strategy: Vice or virtue? *Academy of Management Review*, *11*, 408–427.

Browning, L., Beyer, J., & Shetler, J. (1995). Building cooperation in a competitive industry: SEMATECH and the semiconductor industry. *Academy of Management Journal*, *38*, 113–151.

Caves, R., & Porter, M. (1977). From entry barriers to mobility barriers: Conjectural decisions and contrived deterrence to new competition. *Quarterly Journal of Economics*, *91*, 241–261.

Chandler, A. (1977). *The visible hand*. Cambridge, MA: Belknap.

Cool, K., & Dierickx, I. (1993). Rivalry, strategic groups, and firm profitability. *Strategic Management Journal*, *14*, 47–59.

Cool, K., & Schendel, D. (1987). Strategic group formation and performance: The case of the U.S. pharmaceutical industry, 1963–1982. *Management Science*, *33*, 1102–1124.

Dranove, D., Peteraf, M., & Shanley, M. (1998). Do strategic groups exist? An economic framework for analysis. *Strategic Management Journal*, *19*, 1029–1044.

Dranove, D., & Shanley, M. (1995). Cost reductions or reputation enhancement as motives for mergers: The logic of multihospital systems. *Strategic Management Journal*, *16*, 55–74.

Ferguson, T., Deephouse, D., & Ferguson, W. (2000). Do strategic groups differ in reputation? *Strategic Management Journal*, *21*, 1195–1214.

Ferguson, T., & Ketchen, D. (1999). Organizational configurations and performance: The role of statistical power in extant research. *Strategic Management Journal*, *20*, 385–395.

Garcia-Point, C., & Nohria, N. (2002). Local versus global mimetism: The dynamics of alliance formation in the automobile industry. *Strategic Management Journal*, *23*, 307–321.

Granovetter, M. (1998). Coase revisited: Business groups in the modern economy. In: G. Dosi, D. Teece & J. Chytry (Eds), *Technology, Organization, and Competitiveness: Perspectives on industrial and corporate change* (pp. 67–103). Oxford, UK: Oxford University Press.

Gulati, R. (1995). Does familiarity breed trust? The implications of repeated ties for contractual choice in alliances. *Academy of Management Journal*, *38*, 85–112.

Hatten, K., & Hatten, M. (1987). Strategic groups, asymmetrical mobility barriers, and contestability. *Strategic Management Journal*, *8*, 329–342.

Hayes, S., Spence, A., & Marks, D. (1983). *Competition in the investment banking industry*. Cambridge, MA: Harvard University Press.

Houthoofd, N., & Heene, A. (2002). The quest for strategic groups: Overview, and suggestions for future research. Working Paper, Universiteit Gent.

Human, S., & Provan, K. (1997). An emergent theory of structure and outcomes in small-firm strategic manufacturing networks. *Academy of Management Journal*, *40*, 368–403.

Hunt, M. (1972). *Competition in the major home appliance industry, 1960–1970*. Unpublished doctoral dissertation, Harvard University.

Jobber, D., & Lucas, G. (2000). The modified Tichy TPC framework for pattern matching development in historical case study research. *Strategic Management Journal*, *21*, 865–874.

Ketchen, D., Thomas, J., & Snow, C. (1993). Organizational configurations and performance: A comparison of theoretical approaches. *Academy of Management Journal*, *36*, 1278–1313.

Khanna, T., & Rivkin, J. (2001). Estimating the performance effects of business groups in emerging markets. *Strategic Management Journal*, *22*, 45–74.

Madhavan, R., Koka, B., & Prescott, J. (1998). Networks in transition: How industry events (re)shape interfirm relationships. *Strategic Management Journal*, *19*, 439–460.

Marsh, S. (1998). Creating barriers for foreign competitors: A study of the impact of anti-dumping actions on the performance of U.S. firms. *Strategic Management Journal*, *19*, 25–38.

Mascarenhas, B. (1989). Strategic group dynamics. *Academy of Management Journal*, *32*, 333–352.

McGahan, A., & Porter, M. (1997). How much does industry matter, really? *Strategic Management Journal*, *18*(Summer Special Issue), 15–30.

McGee, J., & Thomas, H. (1986). Strategic groups: Theory, research and taxonomy. *Strategic Management Journal*, *7*, 141–160.

McNamara, G., Deephouse, D., & Luce, R. (2003). Competitive positioning within and across a strategic group structure: The performance of core, secondary, and solitary firms. *Strategic Management Journal*, *24*, 161–181.

Micklethwait, J., & Woolridge, A. (2003). *The company: A short history of a revolutionary idea*. New York: Random House (Modern Library).

Nair, A., & Filer, L. (2003). Cointegration of firm strategies within groups: A long-run analysis of firm behavior in the Japanese steel industry. *Strategic Management Journal*, *24*, 145–159.

Nair, A., & Kotha, S. (2001). Does group membership matter? Evidence from the Japanese steel industry. *Strategic Management Journal*, *22*, 221–235.

Nath, D., & Gruca, T. (1997). Convergence across alternative methods for forming strategic groups. *Strategic Management Journal*, *18*, 745–760.

Osborn, R., & Hagedoorn, J. (1997). The institutionalization and revolutionary dynamics of interorganizational alliances and networks. *Academy of Management Journal*, *40*, 261–278.

Oster, S. (1982). Intraindustry structure and the ease of strategic change. *Review of Economics and Statistics*, *64*, 376–383.

Pegels, C., Song, Y., & Yong, B. (2000). Management heterogeneity, competitive interaction groups, and firm performance. *Strategic Management Journal*, *21*, 911–923.

Peteraf, M. (1993). Intraindustry structure and response toward rivals. *Journal of Managerial and Decision Economics*, *14*, 519–528.

Peteraf, M., & Shanley, M. (1997). Getting to know you: A theory of strategic group identity. *Strategic Management Journal*, *18*(Special Summer Issue), 165–186.

Porter, M. (1976). *Interbrand choice, strategy and bilateral market power*. Cambridge, MA: Harvard University Press.

Porter, M. (1979). The structure within industries and companies' performance. *Review of Economics and Statistics*, *61*, 214–227.

Punj, G., & Stewart, D. (1983). Cluster analysis in marketing research: Review and suggestions for application. *Journal of Marketing Research*, *20*, 134–148.

Reger, R., & Huff, A. (1993). Strategic groups: A cognitive perspective. *Strategic Management Journal*, *14*, 103–123.

Rowley, T., Baum, J., Shipilov, A., Greeve, H., & Rao, H. (2004). Competing in groups. *Managerial and Decision Economics*, *25*, 453–472.

Ruefli, T., & Wiggins, R. (2003). Industry, corporate, and segment effects and business performance; A non-parametric approach. *Strategic Management Journal*, *24*, 861–880.

Rumelt, R. (1991). How much does industry matter? *Strategic Management Journal*, *12*, 167–185.

Rumelt, R. P. (1984). Towards a strategic theory of the firm. In: R. B. Lamb (Ed.), *Competitive strategic management* (pp. 556–570). Englewood Cliffs, NJ: Prentice-Hall.

Russell, B. (1998). *The problems of philosophy* (2nd ed.). Oxford, UK: Oxford University Press.

Sakakibara, M. (2002). Formation of R&D consortia: Industry and company effects. *Strategic Management Journal*, *23*, 1033–1050.

Scherer, F., & Ross, D. (1990). *Industrial market structure and economic performance*. Boston, MA: Houghton Mifflin.

Schmalensee, R. (1989). Interindustry studies of structure and performance. In: R. Schmalensee & R. Willig (Eds), *Handbook of industrial organization*, (Vol. 2, pp. 951–1010). Amsterdam: North-Holland.

Scott Morton, F. (1997). Entry and predation: British shipping cartels 1879–1929. *Journal of Economics and Management Strategy*, *6*, 679–724.

Shanley, M., & Peteraf, M. (2004). Vertical group formation: A social process perspective. *Managerial and Decision Economics*, *25*, 473–488.

Short, J., Ketchen, D., Palmer, T., & Hult, G. (2004). *Firm, strategic group, and industry influences on performance*. Unpublished Working Paper.

Spar, D. (1994). *The cooperative edge: The internal politics of international cartels*. Ithaca, NY: Cornell University Press.

Stuart, T. (2000). Interorganizational alliances and the performance of firms: A study of growth and innovation rates in a high-technology industry. *Strategic Management Journal*, *21*, 791–811.

Thomas, H., & Venkatramen, N. (1988). Research on strategic groups: Progress and prognosis. *Journal of Management Studies*, *25*, 537–555.

Tremblay, V. (1985). Strategic groups and the demand for beer. *Journal of Industrial Economics*, *34*, 183–198.

Wiggins, R., & Ruefli, T. (1995). Necessary conditions for the predictive validity of strategic groups: Analysis without reliance on clustering techniques. *Academy of Management Journal*, *38*, 1635–1655.

Zaheer, S., Albert, S., & Zaheer, A. (1999). Time-scales and organizational theory. *Academy of Management Review*, *24*, 725–741.

ARE REAL OPTIONS "REAL"?

Timothy B. Folta

ABSTRACT

I am interested in clarifying the discussion of how researchers might try to isolate real option effects to identify whether managerial decisions are guided by a real option heuristic. If we are to claim that the theory of real options illuminates managerial behavior, then as a field, we must converge on an understanding as to what constitutes a real option effect, and what does not. The discussion centers on hypothesis development, measurement issues, and research methodology.

Understanding how organizational choice affects the direction and performance of organizations is the heart of strategic management. Issues of organizational choice that have been the subject of strategy research include diversification, integration, alliances, R&D investment, capital investment, and multinational organization, among many others. That real option reasoning has begun to contribute to that inquiry is understandable. Strategic choice involves commitment of resources, which under conditions of uncertainty, places the firm at some hazard. Real options theory offers a systematic way of analyzing the relative merits of different investment alternatives in the midst of uncertainty, and a set of testable propositions relating to those merits to attributes of transactions, firms, and the surrounding environment. In effect, real options theory offers strategic

Research Methodology in Strategy and Management
Research Methodology in Strategy and Management, Volume 2, 93–109

ISSN: 1479-8387/doi:10.1016/S1479-8387(05)02005-9

decision-makers a set of normative rules for choosing among investment alternatives when choice is truly strategic. To the extent that investment choices are an important determinant of firm performance, managers would be well advised to heed those rules and to factor real options considerations into their decision-making calculus.

There is one broad methodological concern on which this chapter is focused. I am concerned with *understanding whether managers recognize and act upon the flexibility* inherent in their decision alternatives, even if managers do not precisely and quantitatively value flexibility.[1] More specifically, I am interested in clarifying the discussion of how researchers might try to isolate real option effects to identify whether managerial decisions are guided by a real option heuristic. If we are to claim that the theory of real options illuminates managerial behavior, then as a field, we must converge on an understanding as to what constitutes a *real option effect*, and what does not. My conclusion is that this convergence is currently lacking, leaving research in real options vulnerable to critics not ready to claim that this perspective adds to our understanding of strategic behavior.

REAL OPTIONS AND STRATEGY

The last 20 years have seen an enormous increase in the interest among strategic management scholars in issues pertaining to flexibility. Needless to say, the real options perspective has been at the forefront of that development. Building off the pathbreaking work of Fischer Black, Myron Scholes, and Robert Merton (1973), Stewart Myers provided the logical foundation for analyzing an organization's investment decisions as real options when he emphasized that "many corporate assets, particularly growth opportunities, can be viewed as call options" (Myers, 1977). His seminal article has spurred a plethora of work in financial economics, that seeks to use option pricing techniques to formally value real assets. While Myers is a notable scholar in financial economics at the Wharton School of Management, his ideas resonated with colleagues in the Management Department, including Bruce Kogut, Ed Bowman, and a doctoral student, Dileep Hurry. It was the work of these scholars (Hurry, 1990; Hurry, Miller, & Bowman, 1992; Kogut, 1991) and the book by Avinash Dixit and Robert Pindyck (1994) that opened the floodgates to empirical research in strategic management and economics, which seeks to understand whether real options theory illuminates managerial behavior. The main substantive contribution of real options theory has been to relate the opportunity costs associated with

commitment and the advantages of flexibility associated with incremental investment, to the attributes of an investment alternative "in a discriminating way"; which is to say, in a way that permits the hypotheses about organizational form to be formulated and tested.

The success of the theory in generating testable hypotheses has led to a recent explosion of empirical research. Although a comprehensive review of that literature is beyond the scope of this chapter, a conservative assessment is that the empirical findings have been broadly supportive of real options propositions. Still, empirical investigation into the fruitfulness of real option theory in explaining managerial behavior involves some clear understanding as to the main constructs representing real option effects, and the suitability of the measurement of these constructs. In the remainder of this section, I offer my decidedly biased views on this subject.

In the Beginning

We begin by considering an important question. What is the valuation approach underlying the theories of the firm we adopt in strategic management? Consider, for example, transaction cost theory. Granted, firm managers should choose a governance mode based on the lowest combination of transaction and production cost. However, should we assume that managers are to appraise the relative cash flows using traditional discounted cash flow (DCF) methodology? The resource-based theory offers another example. Penrose (1959) was not explicit in dictating a valuation method, so Rubin (1973), in his mathematical implementation of Penrose's theory of firm growth, explicitly assumes that firms will take a DCF approach to evaluating business opportunities deriving from its slack resources.

The maintained hypothesis underlying neoclassical investment theory is that investment is made when expected performance exceeds some threshold. If we let I^* signify the investment decision, a representative model of the choice between investment and non-investment would be

$$I^*_{ij} = \begin{cases} \text{invest} & \text{if } P_{jt} \geq T_{jt} \\ \text{do not invest} & \text{if } P_{jt} < T_{jt} \end{cases} \tag{1}$$

where P_{jt} represents the expected net present value (NPV) to an investment by firm j at time t, and T_{jt} represents firm j's threshold level of performance required to induce investment. The simple DCF approach is guided by the assumption that T_{jt} is where NPV equals zero, an assumption that incorporates

the opportunity cost of capital.[2] Real options theory revises T_{jt} from the neoclassical level, to give some weight to the opportunity costs and growth opportunities associated with investment. In the midst of uncertainty, it should be emphasized that the size of the revision can be consequential. This suggests that theories of the firm that implicitly assume a DCF valuation approach, do not take real options into account and may be subject to error.

The points I am trying to accentuate are several.

- Valuation approaches underlie every theory of the firm, even if theories are not explicit in specifying the approach.
- It is probably fair to characterize the default method of valuation as traditional discounted cash flow, which emanates from the neoclassical theory of investment.
- Like other valuation approaches, real options theory should be viewed as complementing existing theories of the firm. Let us be clear that real options theory is *not* a theory of the firm. That is, it does not define the conditions that explain the boundaries of the firm.
- Real options theory "does not describe the level of investment *per se*, but identifies the critical threshold required to trigger investment" (Dixit & Pindyck, 1994, p. 421). It was introduced because existing theories of investment inadequately identify the investment threshold – they do not deal adequately with resource commitment in the face of uncertainty. A focus on threshold effects is warranted and overdue.

These points have at least two important implications for how empirical researchers might seek to test whether managers employ a real options heuristic. First, *the assessment of a real option effect requires a comparative assessment of how managerial decisions would differ if traditional discounted cash flow approaches were assumed*. Second, *we should look for evidence of real option effects in investment thresholds*. Each of these implications is discussed below.

UNDERSTANDING THE RELATIVE CONTRIBUTION OF REAL OPTION THEORY

It has long been recognized that models based on a simple DCF decision rule are weakened by their failure to explicitly account for the additional opportunity costs of irreversibility when capital investments are sunk (Abel, 1983; Arrow, 1968). Traditional models of investment calculate the costs of capital

as the product of the weighted average cost of capital and the stock of capital employed. Such an approach fundamentally takes the costs of the capital employed as a period "rental cost" and implicitly assumes that capital can be redeployed to other uses. When capital is substantially sunk (i.e., has a salvage value substantially less than its purchase price), the firm faces additional opportunity costs due to the loss of flexibility that results from "committed," difficult to reverse actions. It has only been with the development of the "real options" approach to investment that the tools necessary for addressing these complications have become available and modeling their implications in a parsimonious manner plausible (Dixit, 1989; Pindyck, 1991). In fact, this approach demonstrates that these opportunity costs are proportional to the level of uncertainty associated with the investment as well as the degree of irreversibility.[3] As uncertainty increases, the value of a flexible strategic position increases relative to a more irreversible one.

Consideration of total uncertainty, rather than just the systematic component of it, is one important feature which distinguishes the real options perspective from traditional investment theory. This perspective recognizes that it is important to distinguish between exogenous and endogenous uncertainty. Exogenous uncertainty is defined as being "largely unaffected by firm actions, and is predominantly resolved over time" (Folta, 1998). The presence of exogenous uncertainty may make it beneficial to delay full commitment and allow for new information to arrive that may affect the desirability or timing of investment. Foregoing such an option creates an opportunity cost that must be included as part of the cost of investment. Endogenous uncertainty "can be decreased by the actions of the firm" (Folta, 1998). When firms can reduce uncertainty by committing resources, it may make sense to do if the opportunity costs associated with commitment are offset by the benefits from learning. This sort of uncertainty combined with an ability to learn from investment activities, encourages incremental investment.

Given the importance of exogenous and endogenous uncertainty, a key question is how might empirical researchers measure and model these variables in a way that allows researchers to distinguish a real option effect. The subsequent subsections attend to this question.

Measurement of Exogenous Uncertainty

There are at least two key challenges to measuring exogenous uncertainty. The first challenge involves choosing an appropriate unit of analysis (i.e., firm, industry, macroenvironment). It seems important that two conditions

be met. First, the measure of uncertainty should be pertinent to the context in which the investment is being considered. For example, Campa (1993) studied the effect of exchange rate uncertainty on foreign entry decisions. Clearly, exchange rate uncertainty is one factor that may bear upon the value of a commitment into a country. Folta and Miller (2002) examined the impact of volatility on a biotechnology stock market index as representative of the technological and regulatory uncertainty in the industry that might bear upon the decision to buy out a biotechnology partner. The second condition is that the measure of uncertainty is exogenous, or nearly exogenous, to the firm. This rules out immediately measures of firm-level uncertainty, because irreversibility has greater consequence in the case of industry-wide uncertainty. As stated by Dixit and Pindyck (1994), "If one steel firm suffers an idiosyncratic negative shock, it can sell its plant to another firm and get fairly good value for it, so the irreversibility is less severe. However, if the whole industry suffers a negative shock, then the resale value of the plant is small and the irreversibility is large" (p. 249). Whether one measures uncertainty at the industry-wide level or the macroeconomic level should bear predominantly on which type of uncertainty has the greatest influence on the irreversibility of the asset in question. There are a multiplicity of sources that account for the randomness in the external industry or macroeconomic environment. Exogenous uncertainty has been represented in a variety of way in empirical studies, including the volatility of demand (Episcopos, 1995; Folta & O'Brien, 2004; Kogut, 1991; Price, 1995), exchange rates (Campa, 1993), inflation (Huizinga, 1993), output prices (Huizinga, 1993), and stock market indices (Episcopos, 1995; Folta, 1998; Folta et al., 2005; Folta & Miller, 2002; Pindyck, 1986).

A second challenge in the measurement of exogenous uncertainty involves the development of a time-varying measure. It has been relatively common for empirical researchers testing for real options effects to measure exogenous uncertainty by calculating the variance of some output or indicator (e.g., stock price, GDP, sales, etc.) over time. This approach has two critical shortcomings. First, it fails to account for the trends in the data, which will increase the measured variance although they may not constitute an element of uncertainty if they are predictable. Second, this approach does not allow for the possibility that the variance may be heteroskedastic (i.e., not constant over time), a characteristic that is typical of many economic time series.

Fortunately, there is a methodology emanating from the pioneering work of Nobel Laureate Robert Engle that reconciles both of these concerns. The conditional variance generated from generalized autoregressive conditional heteroskedasticity (GARCH) models (Bollerslev, 1986), approximate

unique, time-varying estimates of uncertainty. The GARCH model produces an estimate of the conditional variance, which captures the uncertainty that is not predictable about any trend that might exist for each period in the time series. There is some precedent in the real options literature for employing these models to approximate exogenous uncertainty, and they are being increasingly recommended to capture multi-period forecasts of volatility (Campbell, et al., 1997).

Measurement of Endogenous Uncertainty

Less is known about how one might measure endogenous uncertainty, i.e., uncertainty that can be resolved by firm action. We submit, however, that the inverse of endogenous uncertainty is learning. Firms that learn more efficiently, are better able to reduce uncertainty surrounding their opportunities. In this sense, there is tremendous opportunity to link real options theory with the evolutionary theory of the firm. The Cohen and Levinthal (1990) construct of absorptive capacity and Kogut and Zander's (1992) notion of combinant capability might help in understanding when firms are better able to reduce endogenous uncertainty. Firms that are better able to learn from their experiences might be more suitable candidates for incremental investments in growth options.

I imagine some would ask, "What then, would distinguish real options theory from the evolutionary theory of the firm?" It is the predicted interaction effect. Whereas an evolutionary perspective would simply consider the main effect of absorptive capacity on a strategic outcome, a combined real option and evolutionary perspective would look at the interaction between exogenous uncertainty and absorptive capacity.[4] No work that we know of has interacted firm-level measures of learning with exogenous uncertainty. As such, an interaction between exogenous uncertainty and absorptive capacity might enable researchers to isolate real option effects tied to firm learning. For example, in studying whether managers adopt real options heuristics when considering entry into a technical subfield, one approach would be test if the interaction between a firm's absorptive capacity and the subfield's exogenous uncertainty influences the entry decision.

Interactions with Other Variables

As discussed above, discerning real option effects relative to other effects may be effectively done by interacting exogenous uncertainty with other

variables that might bear upon option value. An examination of option pricing models, such as the Black-Scholes model, reveals that volatility, which we equate with exogenous uncertainty, has a main effect on option value, but also interacts with every other variable (exercise price, time to expiration, interest rate, asset value) in the equation. It is, important to also note that in those models, variables other than volatility have no main effect. This suggests that empirical work should focus on exogenous uncertainty, and interactions with exogenous uncertainty to determine the contribution of real options relative to existing theoretical frameworks.

There is some work that illustrates this strategy. For example, Folta et al. (2005) interacted exogenous uncertainty with measures of irreversibility (i.e., exercise price) to isolate the determinants of value in the option to defer entry into new markets. They reasoned that uncertainty will have no effect in the absence of irreversible investment, but that its effect will become more significant with greater levels of irreversibility.

Interactions of Multiple Types of Real Options

One of the challenges of testing for real option effects is that most investment decisions involve multiple real options. For example, Kulatilaka and Perotti (1998) noted that most investments involved a tradeoff between the option to defer and the option to grow. Understanding how these options overlap is not necessarily obvious Trigeorgis (1986). Empirical researchers trying to discern real option effects must evaluate which options to consider and how to model their different effects.

To date, most of the empirical research on real options theory has focused on the option to defer, which ascribes value to avoiding commitment in the face of exogenous uncertainty. Thus, a real options perspective focusing on the option to defer would argue for a negative relationship between exogenous uncertainty and investment. Alternatively, the option to grow gains its value from the possibility that early investment will help the firm to develop a "capability" that will allow it to take better advantage of future growth opportunities in the industry (Kulatilaka & Perotti, 1998). Accordingly, more valuable growth options encourage investment, which suggests a positive relationship between the likelihood of investment and exogenous uncertainty. Folta and O'Brien (2004) tested the conjecture of Kulatilaka and Perotti (1998) that the combination of the two options would lead to a non-monotonic effect of exogenous uncertainty. Their results suggested that uncertainty has a negative effect on entry at low levels

of uncertainty and a positive effect at high levels, where growth options tend to dominate. The success of their endeavor suggests that future work should consider how to model the effect of overlapping options in a single investment decision.

A related issue that needs further illumination is where a firm has a real option portfolio, and option values overlap within the portfolio. Vassolo et al. (2004) considered the fact that investments by pharmaceutical firms in biotechnology firms represent growth options on new technology. To the extent that a firm's options may overlap, additions to (or subtractions from) their option portfolio may not be accurately represented by merely taking the value of the additional single option into account. This work encourages some attention to the effects of real option portfolio overlap, rather than an exclusive focus on isolated real options.

Controlling for Systematic Risk

If we are to provide evidence that real options theory provides additional explanatory power relative to traditional DCF approaches, we should include measures of systematic risk as control variables. Incremental models should add measures of total uncertainty and determine whether there is added explanatory power. It is true that interpretation of the effect between total uncertainty and the dependent variable is confounded by the inclusion of both total uncertainty and systematic risk. Therefore, if there is evidence that the inclusion of total uncertainty improves model fit, then researchers should consider withdrawing the systematic risk variable to better ascertain the relationship between total uncertainty and the dependent variable.

To summarize this section, we believe that efforts to isolate real option effects need to focus on the effect of total uncertainty on strategic decisions, as compared to just the systematic component of uncertainty. Interactions with exogenous uncertainty offer a fascinating opportunity to examine how existing theories of the firm might be revised when managers employ a real options heuristic. Careful measurement of exogenous uncertainty is critical.

REAL OPTION EFFECTS IN INVESTMENT THRESHOLDS

Eq. (1) highlights the importance of thresholds in theories of investment. Real option effects should be noted in the critical threshold required to

trigger investment. Despite the theoretical clarity that the presence of options influences investment decisions through an impact on firm thresholds, empirical verification on this impact remains uncultivated. The most direct test of option theory would be to measure the effects of uncertainty and irreversibility on threshold levels. However, the fact that firm thresholds are not directly observable confounds the ability to test theory (Dixit & Pindyck, 1994). Scholars have sought to "work around" this challenge in a number of ways.

Empirical Precedents

A number of studies have focused on observing discrete investment events, such as entry, and regressing option variables against those events. This approach implicitly recognizes that entry is a function of a firm's expected profits and its unique threshold. If option variables have no effect on expected performance, their impact on entry should illuminate their true impact on the firm's threshold. In this line of inquiry, Campa (1993) examined the effect of exchange rate uncertainty in entry by foreign firms into the U.S., and found that uncertainty had a larger negative effect on entry. He also found that the negative effect uncertainty was most pronounced when entry required more irreversibility. Folta et al. (2005) corroborated these findings when examining the effect of industry uncertainty on entry into new industries by the U.S. firms listed in Compustat. They extended the work of Campa (1993) by providing a more robust test for the moderating effect of irreversibility on uncertainty, including evidence for the effect of firm-level irreversibility. While option theory has the potential to illuminate firm-specific thresholds, these empirical findings are the first to hint that thresholds vary across firms. While providing no direct evidence of a relationship for uncertainty and irreversibility on firm thresholds, this work examining the entry decision suggests that such a relationship exists. It obfuscates, however, the fact that certain variables may effect entry by their influence on expected performance, the threshold for performance, or a combination of the above.

An alternative empirical strategy for investigating the existence of threshold effects in the presence of irreversibility was used by Pindyck and Solimano (1993), and Caballero and Pindyck (1996). They calculate a measure of the marginal profitability of capital, and use the volatility of this series as a proxy for uncertainty, together with its extreme values as an indicator of the threshold at which investment will be triggered. Pindyck and Solimano (1993) used indicators of the threshold to find evidence that countries with

more volatile marginal capital profitability require higher investment thresholds, as predicted by theory. Caballero and Pindyck (1996) used the same approach to examine the differences of investment levels across industries. The problem with this approach is the need to impose a great deal of model structure (e.g., Cobb–Douglas, constant returns technology with a perfectly competitive economy with free entry, etc.) (Carruth et al., 2000). Furthermore, their findings are weakly supportive of theory since the size of the effect of uncertainty is quite small. For example, an increase of 0.05 in the volatility is associated with a 5–15% increase in the threshold. Finally, the thresholds they ascribe do not vary across firms – they are constant within an industry, or a country. While this work does not represent a direct test of the impact of increased uncertainty or irreversibility on investment, it does give consideration to the non-linearities in the investment process that more traditional approaches ignore, as outlined below.

Most empirical studies have sought to validate the predictions of real options theory by testing whether greater uncertainty reduces the level of aggregate (total industry or macroeconomic) or firm investment. For example, Pindyck (1986) demonstrated a negative correlation between the variance of lagged stock market returns and aggregate investment spending in the U.S. Episcopos (1995) found that the level of fixed investment in the U.S. is inversely related to five different types of macroeconomic uncertainty. Carruth et al. (2000) summarized various empirical studies that attempt to correlate aggregate and firm investment with proxy measures of uncertainty. The broad consensus is that there seems to be a significant negative relationship between aggregate uncertainty and investment. Curiously, studies are far less conclusive regarding the relationship between uncertainty and firm levels of investment. While a number of studies report the expected negative relationship between uncertainty and firm investment (Campa, 1993; Guiso & Parigi, 1999; Huizinga, 1993), a number of studies also report weak or no relationship (Campa & Goldberg, 1995; Driver et al., 1996; Leahy & Whited, 1996). The studies trying to validate the usefulness of option theory by examining the impact of uncertainty on investment levels suffer from several limitations. The first limitation is that the theory does "not describe investment *per se*, but rather the critical threshold required to trigger investment" (Dixit & Pindyck, 1994, p. 422). The second limitation is that while several studies have had the firm as the unit of analysis, there have been no attempts to identify firm-level differences in investment thresholds. By assuming that industry investments differ in their degree of irreversibility, they have indirectly focused on industry-level differences. Yet, the relationship between uncertainty and investment ought

to be at least as strongly observed at the level of the firm, since investment thresholds are a firm-level attribute, and aggregation at the industry-level may mask firm-level differences (Carruth et al., 2000). The theory "applies most directly to a firm or similar decision-making unit" (Dixit & Pindyck, 1994, p. 421). Different firms can have different technologies or managerial capabilities, and thus, they may be asymmetrically subjected to exogenous shocks. This implies that firms will have different thresholds. Moreover, historical accidents may leave firms with stocks of resources that position them with different thresholds at which entry is undertaken. Unless firm-level studies control for differences in firm thresholds, the effects of uncertainty may be ambiguous or inconclusive.

To summarize these empirical results, although we know that real options should influence the required threshold level of performance to induce investment, there is a lack of evidence linking these variables directly to the threshold level. Presumably, this lack of evidence is tied to the challenges of isolating the determinants of the unobservable threshold.

A Methodological Approach to Isolating Threshold Effects

Estimation of the model in Eq. (1) presents three methodological challenges: (1) the unobservability of P_{jt} when no entry occurs; (2) the endogenous nature of the entry decision (because entry only occurs when expected performance exceeds the threshold); and (3) the total unobservability of the T_{jt}. Fortunately, a fairly standard econometric technique can resolve these problems. The censored regression (or tobit) with unobserved stochastic thresholds (Maddala, 1983; Nelson, 1977; Smith, 1980) is appropriate when the dependent variable is only observed when it falls above a particular level or threshold, and this threshold varies from observation to observation as a function of some independent variables.[5] Thus, this methodology deals with the challenges identified above. This method has two other advantages over approaches used in prior research. First, it avoids potential problems of self-selection bias (Heckman, 1979). Second, censored regression estimation can identify the magnitude of individual coefficients for both P_{jt} and T_{jt}, and therefore permits tests of hypotheses regarding the relationship of variables to each construct.

The decision to invest or not is based on the comparison of the two latent constructs of expected performance and the threshold,

$$P_{jt} = aX + e \tag{2}$$

$$T_{jt} = \beta Z + u \tag{3}$$

X and Z are vectors of attributes thought to influence P_{jt} and T_{jt}, respectively; α and β are coefficient vectors, and e and u are normally distributed random variables. Even though the threshold, T_{jt}, cannot be observed for any observation, and P_{jt} cannot be observed for non-investment, the full structure of the investment decision can be estimated if we know the selection process and if we can observe data or proxies for the expected returns to acquisitions. Since P_{jt} but not T_{jt} is available, the model becomes

$$I_{jt} = \begin{cases} \text{invest} = aX + e & \text{if } P_{jt} \geq T_{ji} \\ \text{do not invest} = \text{n.a.} & \text{if } P_{jt} < T_{jt} \end{cases} \tag{4,5}$$

Consistent estimates of the coefficients of Eqs. (2) and (3) can be obtained as long as (i) an independent variable in X is not in Z, or (ii) the covariance between e and u is 0.

One critical decision left for the empiricist is in defining the variable P_{jt}. A strict interpretation of model (1) would suggest that estimates of expected performance are appropriate in the case of investment. However, how does one arrive at such a calculation? Alternatives that might proxy for expected performance include

a) *Abnormal returns around an investment event*. This assumes that market response corresponds with the project's expected performance. If, however, the market had already incorporated the returns prior to the event, the appropriateness of the measure is in jeopardy.
b) *Accounting returns subsequent to an investment*. This assumes that actual returns correspond closely with expected NPV, and that returns near the event are proportional to total NPV. This is problematic in instances where investment involves research and development, where returns may occur many years after initial investment.
c) *Investment levels corresponding to the event*. This assumes that levels of investment correspond with expected returns. This seems a reasonable assumption.

CONCLUSION

In this chapter, I have attempted to clarify the key constructs, measurement issues, and methodology that might allow researchers to distinguish real option effects from alternative effects. Table 1 offers a checklist of these issues for empirical researchers trying to isolate real options effects.

Table 1. Checklist for discerning whether real options are "real".

Hypothesis development
√ Do hypotheses include a focus on exogenous uncertainty or on interactions with exogenous uncertainty?
√ Do hypotheses distinguish between threshold effects and investment effects?
√ Has some consideration been given to the effect of different types of options on a single investment alternative?
√ Has some consideration been given to the overlap in a firm's portfolio of options?
Methodological issues
√ Is the measure of exogenous uncertainty pertinent to the context in which the investment is being considered (i.e., does it bear upon the irreversibility of the investment)?
√ Does the measure of exogenous uncertainty control for trends in the data and allow for unique, time-varying estimates (i.e., is an ARCH or GARCH model used to estimate)?
√ Is systematic risk controlled for in the empirical model?
√ Does the operationalization of variables or research design allow one to distinguish between (or control for) the effects of multiple types of real options?
√ Does the operationalization of variables or research design allow one to control for the overlap among a firm's portfolio of real options?
√ Does the empirical model allow one to discern statistically significant effects of a firm's threshold for investment?

A greater focus on investment thresholds and the determinants of these thresholds is warranted. Let us be clear that not all determinants of investment thresholds will be related to real options. We should be careful to use theory in guiding us which effects to test. I have offered my own opinions about where we might turn in this regard.

There remains much to be learned about real options, and how to test for real option effects. For example, the question of what constitutes growth options is not very well understood. The vast majority of empirical work has focused on the option to defer, at the exclusion of growth options. Yet, for strategy researchers, growth options are particularly intriguing. In a review of the literature pertaining to entry into new markets and the scale at which firms enter, Caves (1998) noted that the pattern of empirical evidence invites interpretation in terms of entrants' diverse expectations and real options: entrants holding more positive expectations about their untested capabilities – their costs, or the qualities of their assets – make larger initial commitments. Even if the industry's technology supports a large optimal scale, the less confident entrant might rationally start out small, incurring a unit-cost penalty but limiting its sunk commitment while it gathers evidence on its unknown capability. This characterization of entry using the theory of real options has close parallels to March's (1991) concept of *exploration to*

expand firm capabilities, and is the basis for Matsusaka's (2001) award winning theoretical paper on diversification in the *Journal of Business*. Clearly, the evolutionary theory of the firm and the resource based view may offer insight into the determinants of growth options. The focus here should be on firm-level effects, which lie in strong contrast to the majority of empirical work on real options.

Another fruitful area of research is exit. This is an understudied phenomenon in strategy, and real options theory offers one perspective on what might influence exit. In fact, in the context of exit, it may be much easier to test the threshold model elaborated upon above. One might approximate P_{jt} with a firm's actual performance.

In conclusion, let me emphasize that this work represents the obvious bias of this researcher. My hope is that I do not constrain work in this area. Rather, the hope is that the field converges on some fundamental approaches to examining the fruitfulness of this perspective and its potential to complement existing theories of the firm.

NOTES

1. Survey evidence has indicated that even though most managers report that real options are important in influencing investment decisions, few firms have formal systems for valuing them (Busby & Pitts, 1997).
2. This assumes that investment opportunities are not mutually exclusive. If investment opportunities are mutually exclusive, the firm is advised to invest in the project with the highest NPV, as long as it exceeds zero. It is relatively common in practice for managers to assume that opportunities are not mutually exclusive.
3. In employing the term "uncertainty," we are referring to exogenous uncertainty, or the volatility of the stochastic process determining the returns from investment, consistent with Dixit and Pindyck (1994).
4. Since the evolutionary theory presumably uses traditional DCF approaches to valuation, it would not, by itself consider total uncertainty to be consequential.
5. This method has been used to estimate the determinants of actual and reservation wages in labor supply applications (Nelson, 1977), market transaction costs, and internal organizational costs in studying decisions around organizational forms (Masten et al., 1991), and entrepreneur's performance and performance threshold in relation to the exit decision (Gimeno et al., 1997).

ACKNOWLEDGMENTS

I am grateful to Jonathan O'Brien and Doug Johnson, whose thoughtful discussions on previous projects where helpful in crafting this chapter.

REFERENCES

Abel, A. B. (1983). Optimal investment under uncertainty. *American Economic Review*, *73*(1), 228–233.

Arrow, K. J. (1968). Optimal capital policy with irreversible investment. In: J. N. Wolfe (Ed.), *Value, capital and growth, essays in honor of Sir John Hicks*. Edinburgh, Scotland: Edinburgh University Press.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*, 307–327.

Busby, J. S., & Pitts, C. G. C. (1997). Real options in practice: An exploratory survey of how finance officers deal with flexibility in capital appraisal. *Management Accounting Research*, *8*, 169–186.

Caballero, R. J., & Pindyck, R. S. (1996). Uncertainty, investment, and industry evolution. *International Economic Review*, *37*(3), 641–662.

Campa, J. M. (1993). Entry by foreign firms in the United-States under exchange-rate uncertainty. *Review of Economics and Statistics*, *75*(4), 614–622.

Campa, J., & Goldberg, L. S. (1995). Investment in manufacturing, exchange-rates and external exposure. *Journal of International Economics*, *38*(3–4), 297–320.

Campbell, J. Y., Lo, A. W., & MacKinlay, A. C. (1997). *The econometrics of financial markets*. Princeton, NJ: Princeton University Press.

Carruth, A., Dickerson, A., & Henley, A. (2000). What do we know about investment under uncertainty? *Journal of Economic Surveys*, *14*(2), 119–153.

Caves, R. E. (1998). Industrial organization and new findings on the turnover and mobility of firms. *Journal of Economic Literature*, *36*(4), 1947–1982.

Dixit, A. K. (1989). Entry and exit decisions under uncertainty. *Journal of Political Economy*, *97*, 620–638.

Dixit, A. K., & Pindyck, R. S. (1994). *Investment under uncertainty*. Princeton, NJ: Princeton University Press.

Driver, C., Yip, P., & Dakhil, N. (1996). Large company capital formation and effects of market share turbulence: Micro-data evidence from the PIMS database. *Applied Economics*, *28*(6), 641–651.

Episcopos, A. (1995). Evidence on the relationship between uncertainty and irreversible investment. *Quarterly Review of Economics and Finance*, *35*(1), 41–52.

Folta, T. B. (1998). Governance and uncertainty: The trade-off between administrative control and commitment. *Strategic Management Journal*, *19*(11), 1007–1028.

Folta, T. B., & O'Brien, J. (2004). Entry in the presence of dueling options. *Strategic Management Journal*, *25*(2), 121–138.

Folta, T. B., & Miller, K. D. (2002). Real options in equity partnerships. *Strategic Management Journal*, *23*(1), 77–88.

Folta, T. B., Johnson, D. R., & O'Brien, J. (2005). *Uncertainty and the likelihood of entry: An empirical assessment of the moderating role of irreversibility*. West Lafayette, IN: Purdue University.

Gimeno, J., Folta, T. B., Cooper, A. C., & Woo, C. Y. (1997). Survival of the fittest? Entrepreneurial human capital and the persistence of underperforming firms. *Administrative Science Quarterly*, *42*(4), 750–783.

Guiso, L., & Parigi, G. (1999). Investment and demand uncertainty. *Quarterly Journal of Economics*, *114*(1), 185–227.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, *47*, 153–161.
Huizinga, J. (1993). Inflation uncertainty, relative price uncertainty, and investment in United-States manufacturing. *Journal of Money Credit and Banking*, *25*(3), 521–549.
Hurry, D. (1990). *A strategic options framework for multinational exploration and decision-making*. Philadelphia: University of Pennsylvania.
Hurry, D., Miller, A. T., & Bowman, E. H. (1992). Calls on high-technology – Japanese exploration of venture capital investments in the United-States. *Strategic Management Journal*, *13*(2), 85–101.
Kogut, B. (1991). Joint ventures and the option to expand and acquire. *Management Science*, *37*(1), 19–33.
Kogut, B., & Zander, U. (1992). Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science*, *3*(3), 383–397.
Kulatilaka, N., & Perotti, E. C. (1998). Strategic growth options. *Management Science*, *44*(8), 1021–1031.
Leahy, J. V., & Whited, T. M. (1996). The effect of uncertainty on invesmtent: Some stylized facts. *Journal of Money Credit and Banking*, *28*(1), 64–83.
Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics*. Cambridge, UK: Cambridge University Press.
March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science*, *2*(1), 71–87.
Masten, S. E., Meehan, J. W., & Snyder, E. A. (1991). The costs of organization. *Journal of Law Economics & Organization*, *7*(1), 1–25.
Matsusaka, J. G. (2001). Corporate diversification, value maximization, and organizational capabilities. *The Journal of Business*, *74*(3), 409–431.
Merton, R. C. (1973). Theory of rational option pricing. *Bell Journal of Economics and Management Science*, *4*, 141–183.
Myers, S. C. (1977). Determinants of corporate borrowing. *Journal of Financial Economics*, *5*(2), 147–175.
Nelson, F. D. (1977). Censored regression models with unobserved stochastic censoring thresholds. *Journal of Econometrics*, *6*, 309–327.
Penrose, E. T. (1959). *The Theory of Growth of the Firm*. New York: John Wiley.
Pindyck, R. S. (1986). Capital risk and models of investment behaviour. Sloan School of Management (working paper).
Pindyck, R. S. (1991). Irreversibility, uncertainty, and investment. *Journal of Economic Literature*, *29*(3), 1110–1148.
Pindyck, R. S., & Solimano, A. (1993). Economic instability and aggregate investment. *NBER Macroeconomic Annual*, 259–303.
Price, S. (1995). Aggregate uncertainty, capacity utilization and manufacturing investment. *Applied Economics*, *27*(2), 147–154.
Rubin, P. H. (1973). The expansion of firms. *Journal of Political Economy*, *81*, 936–949.
Smith, J. P. (1980). *Female labor supply: Theory and estimation*. Princeton, NJ: Princeton University Press.
Trigeorgis, L.G. (1986). Valuing real investment opportunities: An options approach to strategic capital budgeting. Doctoral Thesis, Harvard University.
Vassolo, R. S., Anand, J., & Folta, T. B. (2004). Non-additivity in portfolios of exploration activities: A real options based analysis of equity alliances in biotechnology. *Strategic Management Journal*, *25*(11), 1045–1061.

THEORY AND METHODOLOGY IN ENTREPRENEURSHIP RESEARCH

R. Duane Ireland, Justin W. Webb and Joseph E. Coombs

ABSTRACT

Entrepreneurship remains a young scholarly discipline characterized by low paradigmatic development. Herein, we discuss theoretical and methodological issues associated with this rapidly emerging yet still developing research area. We argue that theory and methodology are symbiotic components of research and should develop concurrently in order to support the evolution of a paradigm for entrepreneurship research. Further, we posit that effective growth of entrepreneurship research will occur as a result of appropriately extending theory and methods from other scholarly disciplines as well as from theoretical and methodological innovations that are unique to entrepreneurship. Based on the positions taken in this chapter, we also advance recommendations for scholars to consider as work is completed to develop a systematic body of knowledge about entrepreneurship.

INTRODUCTION

The amount of entrepreneurship research scholars are completing continues to dramatically increase, yet debate remains about the core research

Research Methodology in Strategy and Management
Research Methodology in Strategy and Management, Volume 2, 111–141

ISSN: 1479-8387/doi:10.1016/S1479-8387(05)02006-0

questions suggested by the phenomenon itself, as well as how to define the entrepreneurship construct (Gartner, 1990; Shane & Venkataraman, 2000). Researchers define the entrepreneurship phenomenon differently and often use different terms interchangeably (Zahra, Jennings, & Kuratko, 1999). Sharma and Chrisman (1999) integrated various definitions of entrepreneurship (and related terms such as corporate entrepreneurship, strategic renewal, and so forth) in an attempt to bring definitional clarity to many of the field's constructs. Nevertheless, the continuing lack of definitional clarity highlights the status of entrepreneurship as an emerging academic discipline.

The increasing scholarly interest in entrepreneurship is highlighted by the establishment of journals such as *Entrepreneurship Theory & Practice* and *Journal of Business Venturing*. More recently, the formation of the *Journal of International Entrepreneurship* highlights the continuing importance of entrepreneurial activities and guiding scholarship in a global context, and the more frequent inclusion of entrepreneurship-related articles in well-established and highly respected journals (i.e., *Academy of Management Review* and *Strategic Management Journal*) demonstrates the interest of scholars working in different disciplines to examine significant entrepreneurship questions.

Although the progress of entrepreneurship research is clearly promising, there is certainly room for additional theoretical and methodological improvements. As a relatively young discipline, entrepreneurship is characterized by low levels of paradigm development (Kuhn, 1996). Kuhn (1996, pp. 47–48) suggests that less developed paradigms are "regularly marked by frequent and deep debates over legitimate methods, problems, and standards of solution." The relatively low level of development of the entrepreneurship paradigm is perhaps not surprising, given that the field is quite young. As noted by Cooper (2003, p. 5), "Most of the journals we read and most of the conferences we attend were started in the last 20 to 25 years, a little more than half of a professional's working life. Ours indeed is a young field." When working in a field with low paradigm development, scholars must be concerned about the significance of the contributions their work is making to a discipline (Rynes, 2002). Research that makes both theoretical and empirical contributions facilitates a paradigm's development (Bergh, 2003) and leads to scholars having the theoretical insights and methodological tools required to examine significant research questions (Low & MacMillan, 1988; Sexton, 1988).

One of the purposes of this chapter is to describe theory-related issues that are affecting entrepreneurship research. To do this, we first consider some concerns regarding the lack of a clearly articulated theory or theories for entrepreneurship researchers to use as sources for developing significant

entrepreneurship research questions that can be examined with appropriate tests. The concern about the lack of a clear, unique theory of entrepreneurship as a driving force of important empirical work has been mentioned by researchers as a critical issue affecting the quality of entrepreneurship work (Shane & Venkataraman, 2000). Driving this concern is the argument that instead of strong empirical work that is a product of carefully structured theoretical arguments, an appreciable amount of the extant findings reported in the entrepreneurship literature are drawn from case-oriented, anecdotal, and topic-driven research efforts. A common problem researchers encounter with these types of results is the inability to use previously reported findings to draw strong causal attributions (Zahra et al., 1999). In our discussion of theory as a vital aspect of empirical work that is completed by using appropriate methods, we also mention how theory differs from other perspectives such as organizational narratives. Following this brief discussion are descriptions of theories that we believe are relevant for the purpose of specifying important and testable hypotheses regarding entrepreneurship-related phenomena.

A second purpose of this chapter is to describe methods being used to examine entrepreneurship research questions. This description speaks to constraints affecting researchers' decisions about methods they will use to complete their work. For example, scholars interested in studying issues associated with either firm or industry birth often deal with very small populations and samples. Results from statistical tests in these instances are often ambiguous and difficult to interpret. In other instances, researchers interested in studying relationships longitudinally may be prevented from doing so because of the high failure rate of entrepreneurial ventures (hereafter, we use the terms entrepreneurial ventures and entrepreneurial firms interchangeably). Relatedly, because of this high failure rate, researchers must deal with the possibility of survivor bias affecting the quality of their findings. In this section, we deal with four major areas of methodological concerns that we believe affect entrepreneurship research. We also offer recommendations for researchers to consider in dealing with these concerns.

It is perhaps a bit unusual to first deal with theory as the foundation for studying the methods being used by researchers interested in studying particular phenomena. Guiding our decision in this respect is the accepted knowledge that richly developed hypotheses, which are based on careful examinations of one or more theories, are critical to completing important empirical research. Moreover, strong, theory-based arguments establish a foundation for future empirical research (which can include replications and extensions), while empirical studies with strong methodology are the source researchers use to confirm or reject theoretical arguments and to extend

research in new directions. We now turn our attention to theory's role in the design and completion of high-quality entrepreneurship research.

THEORY AND ENTREPRENEURSHIP RESEARCH

Theory development's importance to scholarly research is unquestioned. However, precisely what constitutes theory and a theoretical contribution is less certain (Bacharach, 1989; Sutton & Staw, 1995; Whetten, 1989). Complicating efforts to develop theory and tease out theoretical contributions is the high level of disagreement about theory and methods that often characterize fields with low levels of paradigm development (Hitt, Boyd & Li, 2004). The inability to specify a theory and theoretical contributions captures the current state of research in the entrepreneurship discipline.

Even though a relatively large portion of certainly the early entrepreneurship research lacked effective theory development and subsequently, appropriate empirical tests, the results from these efforts are, nonetheless, valuable for the entrepreneurship field. Indeed, consistent with the nature of growth in the sophistication and quality of research in many other disciplines, the results from primarily case-oriented, anecdotal, and topic-driven work reflect interest in examining a particular phenomenon (entrepreneurship in this instance). More importantly, some of the results from these early efforts provide fundamental insights about the domain and practice of entrepreneurship that in turn become the foundation or backdrop for subsequent theory-building research activities (Sutton & Staw, 1995) as well as the methods useful for testing more richly developed hypotheses. At a certain point, however, research needs to transition to a focus of establishing a systematic body of knowledge by integrating and extending the underlying core ideas noted in this nontheory research. For entrepreneurship, a number of scholars agree that the complexity of the phenomenon and quantity of existing research suggest a current need for theory development.

We address a number of theory issues in this section of the chapter. First, we define theory and theory's importance to organizational scholarship. We then describe theories associated with particular disciplines (e.g., strategic management, sociology, psychology, and economics) that we believe have the potential to inform theory development for designing entrepreneurship research. Finally, we offer recommendations and directions that researchers can follow to develop one or more theories of entrepreneurship. This work is vital, in that arguably, a scholarly discipline cannot stand on its own without an epistemological foundation comprised of its own distinct theories for

predicting and explaining empirical phenomena (Busenitz, West, Shepherd, Nelson, Chandler, & Zacharakis, 2003; Shane & Venkataraman, 2000).

What is Theory?

A theory is "an ordered set of assertions about a generic behavior or structure assumed to hold throughout a significantly broad range of specific instances" (Weick, 1989, p. 517, citing Sutherland, 1975, p. 9). All "good" or effective theory is driven by logic (Wacker, 2004). Drawing from a base of logic, theory explicates the how, why, and when of relationships existing within a system of constructs and variables (Bacharach, 1989; Weick, 1989; Whetten, 1989). Theory explains the causal links between one set of phenomena and at least another set of phenomena (Blalock, 1979). Theoretical links demarcate relationships and the dynamics between or among them. Theory is used to describe how changes in one set of relational objects create changes in other relational objects within the theoretical boundaries. Moreover, theory explains why changes between relational objects occur and specifies any factor that may moderate or mediate the relationship. Strong theory not only identifies causal relationships but also the magnitude of these relationships. Theory should be coherent enough and parameterized within a specific time and space to allow falsifiability, and it should provide usefulness by both enabling prediction and explanation (Bacharach, 1989). As Hrebiniak and Joyce (2001, p. 612) note, "The usefulness of any (theoretically derived) model would be increased to the extent that it accounts for significant amounts of criterion variance with a minimum number of predictors." Theory that incorporates a large number of variables with many having minimal effects becomes cognitively unwieldy and methodologically challenging. Finally, it is important to note that explicit assumptions facilitate the researcher's efforts to create boundaries to the applicability of a theory besides highlighting the underlying basis of what must be true for a theory's validity (Bacharach, 1989; Dubin, 1969).

A competing view of theory contends that a scholarly field benefits more when its definition is expanded to include devices of enlightenment (DiMaggio, 1995), narrative analyses (DiMaggio, 1995; Weick, 1999), and other approaches for general theory approximation and development (Weick, 1995). Devices of enlightenment are ideas aimed at disrupting the current mode of thinking and injecting new insights into established theory (DiMaggio, 1995). Weick (1999) argues that the ability to eloquently communicate narrative analyses can often illustrate processes and their causal

linkages as descriptively as can theory. Moreover, theory approximations, such as post-factum interpretations of empirical results or simple proposition developments, can facilitate scholars' efforts to understand and empirically deal with theoretical intricacies (Weick, 1995). Although each of these methods or approaches used to communicate research results has the potential to uniquely advance organizational theory, explicit links between results and theory are the most effective path to theory development.

Another competing view of theory development argues that theory used in organizational research should be accepted on an almost conditional ground on which extensions and criticism allow the theory to be more fully developed over time on the bases of how, why, and when (Jones, 2001; Weick, 1995). Schumpeter's (1934) view of creative destruction is an example of a theoretical base that has developed over a substantial period of time. In his original conceptualization of creative destruction, Schumpeter (1934) argued that innovation creates market disequilibrium. An entrepreneurial firm disrupts market equilibrium through radical innovation. These actions threaten rival firms and force them to respond; and, the set of actions and reactions create an ever-escalating game of competition (Grimm & Smith, 1997) that encourages innovation while continuing to disrupt the status quo. One of the rich arguments Kirzner (1979) advances in his body of work extended parts of Schumpeter's views. Kirzner (1979) is more concerned with the discovery of market opportunities while Schumpeter (1934) is relatively more concerned with the exploitation of opportunities. From a theoretical perspective, additional work is warranted. For example, due to changing competitive landscapes, further research may enhance our understanding of the sequence of creative, yet destructive events that occur when radical innovation merges multiple technologies and industries.

Theoretical contributions include extensions and/or criticisms that build upon previously developed theories. Theoretical contributions, which are a product of researchers' abilities to establish an *explicit* relationship between the results of their study and grounded theory, permit the continuing evolution of theory in ways that facilitate scholars' efforts to advance a field's knowledge. Thus, entrepreneurship researchers interested in using other theories must establish clear links between their findings and the grounded theory they have used to specify testable relationships. The reason for this is that a theoretical contribution must explain how, why, and when specific phenomena occur as they do and how a set of particular findings and propositions fill a theoretical gap or oppose contemporary beliefs. References, data, variables, diagrams, and hypotheses do not constitute theory or a theoretical contribution (Bacharach, 1989; Sutton & Staw, 1995). The function

of these aspects of scholarly research is to help scholars frame and complete their work in ways that will allow them to describe how their findings contribute to theory and to the use of more appropriate methodologies.

Recent work by Combs, Ketchen, and Hoover (2004) demonstrates the relationship among the effective use of grounded theory to form testable relationships and the subsequent development of theory based on results obtained by using appropriate methods. These researchers used agency theory and resource scarcity theory to predict performance differences among franchising strategic groups. In their work, Combs et al. (2004) clearly explain how the two theories inform the study of important research questions as well as the specification of the tested hypotheses. The following statements speak to the Combs et al. (2004, pp. 880–882) use of theory: "Resource scarcity asserts that firms can use franchising to gain access to critical resources needed to grow quickly and build economies of scale… By using franchising in unfamiliar locations, the firm shifts the burden of understanding local market conditions to franchisees." Furthermore, the researchers used the two theories to specify testable relationships that when examined, would have the potential to fill gaps in the literature regarding the performance of franchising strategic groups. The researchers' use of two theories shows the cumulative nature of theory development as a means of specifying additional, testable relationships.

Combs et al. (2004) found that superior performers (as measured by market-to-book value) responded primarily to agency–theory based arguments rather than the expectations based on resource scarcity theory. While their findings are consistent with agency theory, Combs et al. (2004, p. 891) speculate about the empirical weaknesses of their study that could possibly have led to inconsistencies with the findings with respect to resource scarcity theory: "We were unable to separate the effects of different resource scarcity franchising relationships on performance. Perhaps responding to limited managerial expertise improves performance, whereas responding to capital scarcities do not." Thus, results from the Combs et al. (2004) provide some insights that researchers can use to develop finer-grained theoretical expectations for empirical testing. In a stepwise manner, this type of scholarship advances theory as well as methods used to test theoretically-derived hypotheses.

Organizational Theories and Their Importance for Entrepreneurship Research

A number of organizational theories, including those associated with strategic management, economics, sociology, and psychology, can inform

theory development as well as the use of methods for scholars examining entrepreneurship research questions: Theories from sociology can provide insights about how individuals collectively function within an organization; psychology facilitates forming descriptions of individuals' behaviors; economics presents entrepreneurship researchers with frameworks that can be used to understand efficiency and effectiveness in entrepreneurial ventures; and, strategic management theories can offer insights for how to develop and sustain competitive advantages. Collectively, these insights can inform relationships for entrepreneurship scholars to specify and test. In part, the relationships to be tested could be concerned with differences between entrepreneurial ventures and established organizations. Testable relationships could also be formed, drawing from these fields' grounded theories, regarding the launching of entrepreneurial ventures. In broad terms, integrating theoretical insights from disciplines such as those we cite here has the potential to contribute to intellectually rigorous efforts to form a theory of entrepreneurship (Harrison & Leitch, 1996) or a theory of the firm (Alvarez & Barney, 2004).

Next, we describe several theories with the potential to help entrepreneurship researchers form important and interesting questions, as well as possibly the development of a theory of entrepreneurship. The resource-based view of the firm, network theory, and institutional theory are individually discussed in these contexts.

Resource-based View of the Firm

The resource-based view of the firm (Barney, 1991; Wernerfelt, 1984) states that valuable, rare, inimitable, and nonsubstitutable resource combinations serve as a source of competitive advantage for firms. Resource heterogeneity among and resource immobility across firms are critical assumptions supporting these theoretical expectations. This view of the firm has matured within recent years with both extensions (Amit & Schoemaker, 1993; Sirmon, Hitt, & Ireland, 2005; Teece, Pisano, & Shuen, 1997) and criticism (Priem & Butler, 2001a,b). Some of this work is concerned with dynamically considering resources rather than studying them in a static context. Other contributions (Grant, 1996; Kogut & Zander, 1992) have focused on knowledge resources (e.g., human capital and social capital) as likely sources of sustainable competitive advantages. Knowledge resources' causal ambiguity and social complexity (Barney, 1991; Dierickx & Cool, 1989; Reed & DeFillippi, 1990) are thought to be the reason for these resources to be the foundation of sustainable advantages.

These extensions of the resource-based view yield a valuable source of direction for entrepreneurship research. For example, of the original list of resources Barney (1991) identified, financial capital, human capital, and social capital are now considered to be the entrepreneurial firm's most valuable resources. The value of these resources is that they serve as a platform from which the firm can integrate diverse knowledge stocks to explore for and identify market opportunities that can be exploited through innovation (Ireland, Hitt, & Sirmon, 2003). Research could seek to understand how successful entrepreneurial firms symbiotically manage their limited stocks of these resources to optimize their value (Webb & Ireland, 2004) and to overcome liabilities of newness. Resource symbiosis is created in the firm when resource levels facilitate capturing the optimal value of other resources in the firm. Given entrepreneurial ventures' resource constraints, creating optimal symbiosis implies that the firm is not wasting resources and is better positioned to create value by using its constrained set of resources. Future research could study how these resources are integrated to enhance the firm's entrepreneurial alertness and awareness – attributes that facilitate the recognition of entrepreneurial opportunities. Another interesting entrepreneurship research question that could be grounded in the resource-based view concerns how entrepreneurial ventures integrate their limited resources and skills to form dynamic capabilities as the platform for developing sustainable competitive advantages.

Network Theory

Network theory is an interdisciplinary organizational theory derived from sociology, economics, and strategy research (Busenitz et al., 2003; Low & MacMillan, 1988). The basic premise behind the network theory is that an individual's or an organization's resources and legitimacy are available to it from its network. For resource-deprived entrepreneurial ventures, a social structure in the form of relationships with venture capitalists, business angels, partnering firms and universities, for example, can be a valuable source of both financial capital and knowledge (Batjargal & Liu, 2004; Shane & Stuart, 2002; Steier & Greenwood, 2000; Zaheer & Zaheer, 1997). Nevertheless, while these relationships constitute an additional, possibly more efficient channel of resources, considerable investments are needed to establish and maintain networks, given norms of reciprocity (Adler & Kwon, 2002). Scholarly interest lies in how entrepreneurial firms can service their relationships with their limited resource stocks without losing the source of their competitive advantage (Ireland, Hitt, & Webb, 2005). For example, smaller entrepreneurial ventures often risk the appropriation of their tacit

knowledge by larger dominant firms, which have more available stocks of explicit knowledge and financial capital to share (Alvarez & Barney, 2001). Further interest also lies in whether certain forms of social capital are more valuable to an entrepreneurial firm's overall network. What forms of social capital are valuable to an entrepreneurial venture's ability to identify opportunities? To exploit opportunities? How does an entrepreneurial firm's social structure evolve in value-creating ways? Is there an optimally symbiotic proportion of human capital to social capital in the entrepreneurial firm that enhances the firm's ability to innovate and exploit marketable opportunities, yet maintains a certain value and competitive advantage (Webb & Ireland, 2004)? Do resource-constrained entrepreneurial ventures benefit more from strong, efficient ties (Coleman, 1988), weak, effective ties (Burt, 1992), or a balance between the two (Uzzi, 1997)?

Institutional Theory

Institutional theory is yet another valuable interdisciplinary organizational theory with potential applications to inform entrepreneurship research. The basic tenet of institutional theory is that the firm must conform to standards within its industry to gain legitimacy and power (DiMaggio & Powell, 1983). The firm achieves legitimacy by becoming increasingly similar in resources, structure, and strategy to other firms in the industry (DiMaggio & Powell, 1983). There are three generally recognized forces driving this conformity – coercive isomorphism, normative isomorphism, and mimetic isomorphism (DiMaggio & Powell, 1983). Coercive isomorphism is caused by governmental, cultural, or competitive pressures; normative isomorphism stems from the socialization of members of a common occupation that define a standard for their work methods and conditions; and mimetic isomorphism originates from firm actions taken to imitate or benchmark practices from other organizations in overcoming environmental uncertainty (DiMaggio & Powell, 1983). Isomorphic decisions do not necessarily stem from an organization's attempt to increase efficiency, but rather, are driven in the pursuit of legitimacy and power (DiMaggio & Powell, 1983). For example, the diffusion of ISO 9000 quality certificates within different countries has been linked strongly to each of these isomorphic forces as organizations seek legitimacy for their practices (Guler, Guillen, & Macpherson, 2002). However, must resource-constrained entrepreneurial firms give in to these isomorphic forces and expend what little resources they own or control to gain legitimacy, or can they find ways to circumvent these forces without material harm to their performance? If resource-constrained entrepreneurial firms lack legitimacy but own valuable tacit knowledge, how do they gain power

or create legitimacy within the eyes of established firms? Given that institutional forces drive entrepreneurial firms to utilize similar resources, structures, and strategies as other firms in their industry, yet some aspects of organizations may be more susceptible to these forces than others (Slack & Hinings, 1994), can entrepreneurial firms maintain their unique source of competitive advantage while gaining legitimacy by knowingly filtering certain institutional forces? Another area of interest is how changes in homogeneous environments occur and how firms adapt (Kondra & Hinings, 1998). Do these changes create entrepreneurial opportunities? If so, how can entrepreneurial firms adapt to gain the capability required to identify and then exploit them?

Possible Theory Developments within Entrepreneurship

Despite the complexities that can be part and parcel of developing an organizational theory, the idea behind the theory does not necessarily have to be complex and nonobvious (Weick, 1989). Rather, scholars' knowledge and expertise in their particular fields are the sources from which theories may evolve. The objective of theory development should be an acceptance, rather than a rejection, of an elucidated (how, why, and when) logic, thereby creating a common ground from which empirical and conceptual studies can stem with a consistent basis for further theoretical developments through extensions and criticism. The intent of these secondary contributions is to advance the underlying logic into areas that are not so straightforward and to enhance our understanding of the organization. Over time, scholars' logic can mature into a conceptual base of knowledge that advances our understanding of the organization and our ability to communicate that understanding to others.

There are opportunities to develop theoretical frameworks that are unique to the entrepreneurship field. However, the fragmented nature and diverse scope of previous entrepreneurship research (Gartner, 2001; Shane & Venkataraman, 2000) suggests that efforts to do this will be complex and difficult (Amit, Glosten & Muller, 1993). This complexity and difficulty calls for scholars to initially focus on integrating specific concepts of the entrepreneurial process to establish a common framework from which future theoretical contributions can be derived. For example, Shane and Venkataraman (2000) suggest two possible options for theory development in terms of entrepreneurship: (1) entrepreneurs and (2) entrepreneurial opportunities and the processes of recognizing and exploiting them. Scholars have

attempted to identify basic attributes common to entrepreneurs for a number of decades. Research has utilized mainly a trait- or behavioral-based approach for distinguishing entrepreneurs from non-entrepreneurs (Chrisman & Kellermanns, 2005; Stevenson & Jarillo, 1990). The salience of the context surrounding the entrepreneur and promoting entrepreneurial action created difficulties for the trait-based approach, and research has evolved to identify behaviors to understand how entrepreneurs explore and exploit opportunities (Chrisman & Kellermanns, 2005). Further research has studied additional personal aspects of entrepreneurs, including networks, venture/management experience, and education (Chrisman & Kellermanns, 2005).

Psychological and sociological theories likely will remain applicable (Martinelli, 1994) in explaining the cognitive and behavioral complexity of the entrepreneur as well as environmental factors that affect the development and decision-making styles of the entrepreneur, such as culture and munificence. While there may not exist a single type of entrepreneur, entrepreneurial scholars could apply the logic associated with differentiating strategic decision-making styles and modes (Hart & Banbury, 1994; Hart, 1992; Mintzberg, 1973). If such a framework were formed for a "theory of the entrepreneur," further empirical and conceptual studies could analyze the strengths and weaknesses of certain decision-making modes, certain behaviors and their tendency to lead to entrepreneurial failure, and the cultural effect(s) on certain entrepreneur types.

Scholarly work has been completed toward forming a theory of entrepreneurial opportunity. Eckhardt and Shane (2003, p. 336) define entrepreneurial opportunities as "situations in which new goods, services, raw materials, markets, and organizing methods can be introduced through the formation of new means, ends, or means–ends relationships." Because people's desires, values, experiences, and general knowledge of given contexts differ, individuals view situations as different opportunities or as no opportunity at all (Venkataraman & Harting, 2005). Therefore, opportunity identification depends on both the entrepreneur and the context of the opportunity (Dew, Velamuri & Venkataraman, 2004; Shane, 2000; Venkataraman & Harting, 2005). Furthermore, the opportunity context can be viewed from the perspective of the locus of change, the initiator of change, or the source of opportunity (Eckhardt & Shane, 2003).

Contributions from strategy, sociology, and economics theories could be used to facilitate continuing development of a theory of entrepreneurial opportunity. Additional interest may lie in the conscious and subconscious processes used by an entrepreneur or an entrepreneurial venture to recognize an opportunity and how resources could be managed to support the

bisociative processes (Koestler, 1964) underlying this recognition. Bisociative processes are ones through which two unrelated sets of information and resources are combined (Smith & Di Gregorio, 2002). An opportunity theory likely should also involve an explanation for determining *ex ante* the value of an opportunity as well as the optimal timing for exploiting it. The frameworks provided for entrepreneurial theory development above are in no way intended to be comprehensive. Rather, they solely offer thoughts on possible directions for future theory development in entrepreneurship research.

A preliminary step within theory development should be the differentiation among various definitions for specific entrepreneurship types, providing multiple foci to study from, but each possessing a common definitional grounding. The lack of a commonly agreed upon or accepted definition of entrepreneurship remains an issue affecting the paradigmatic development of entrepreneurship research (Gartner, 1990; Low & MacMillan, 1988; Shane & Venkataraman, 2000; Venkataraman, 1997). This issue creates confusion in theoretical understanding and development as multiple definitions are offered for the same term in some cases, while multiple terms fall under a single definition in others. Scholars are completing work that is intended to result in definitional clarity. Most notably, based on their extensive review of definitions in the literature, Sharma and Chrisman (1999, p. 17) defined entrepreneurship as "acts of organizational creation, renewal, or innovation that occur within or outside an existing organization." The authors further stratified the general definition of entrepreneurship into more specific contexts such as strategic renewal, corporate venturing, independent entrepreneurship, corporate entrepreneurship, and external and internal corporate venturing. However, additional definitions beyond these can be constructed for social, political, and any other number of forms of entrepreneurship. The applicability of any particular definition to a given context will be dependent on the underlying theory, possibly implying that entrepreneurship researchers need to use several theoretical frameworks to capture the complexities of entrepreneurship as a multifaceted phenomenon.

As noted above, establishing a theoretical base from which entrepreneurship scholars can draw to specify testable relationships remains a critical task. Developing a solid theory base would facilitate scholars' efforts to form significant streams of research as part of the pathway to additional legitimacy for entrepreneurship research and to the development of a widely recognized paradigm. In Table 1, we summarize the main recommendations flowing from the points we have addressed in this part of the chapter.

Table 1. Theory Recommendations for Entrepreneurship Research.

	Recommendations
General theory development	Explicate the how, why, and when of relationships among a set of variables Specify necessary assumptions for testable relationships
Definitional issues	Define entrepreneurship in the way in which it is being used Specify the boundaries of the chosen definition of entrepreneurship
Applying theories from other disciplines	Articulate main assertions and assumptions of the theory Discuss applicability to entrepreneurship Discuss how the assertions/assumptions remain the same or change when used to form theory-driven testable relationships dealing with entrepreneurship questions
Entrepreneurship-specific theories	Expand on theories of the entrepreneur and of entrepreneurial opportunity Continue to focus on ways to appropriately develop specific theories about entrepreneurship

Of course, in a symbiotic manner, methodological issues are as important as theory development with respect to the continuing evolution of entrepreneurship research. Next, we discuss methodological issues associated with the entrepreneurship discipline. This discussion is intended to identify issues for entrepreneurship researchers to consider when designing their studies to permit appropriate analyses of their data.

METHODOLOGICAL ISSUES AND ENTREPRENEURSHIP RESEARCH

Although relevant methodologies are often domain specific, some methodological issues are common to all disciplines (see Hitt et al., 2004). Domain specific methodological issues must be effectively addressed for a discipline to grow and develop. In the main, the causes of these issues are the lack of appropriate methodologies to test certain variables and the designs chosen by scholars.

In the following section, we examine the current state of entrepreneurship research. In doing this, we highlight a number of methodological issues, some of which have been associated with entrepreneurship research for

some time. We then distinguish factors that are unique to entrepreneurship research. These factors create difficulties for researchers interested in examining certain entrepreneurship-related phenomena. We offer a set of methodological issues for entrepreneurship scholars to consider when designing their work. We believe that attention to these issues has the potential to facilitate researchers' efforts to successfully design and complete their empirical studies.

Current State of Empirical Entrepreneurship Research

As an emerging discipline, entrepreneurship offers scholars vast opportunities for research. Furthermore, the importance of entrepreneurship to the development of economies across the world signifies the potential significance of such research. However, entrepreneurship scholars committed to designing and executing solid empirical work face methodological challenges. Entrepreneurship is often a highly dynamic process, making it difficult to conduct longitudinal studies on many aspects of entrepreneurship. Further exacerbating longitudinal studies is the high failure rate of entrepreneurial firms, leading to many time-consuming and expensive, yet futile research efforts. Also methodologically challenging is the fact that entrepreneurship encompasses the newest of new technologies and markets. Given this novelty, some industries comprise only a limited population of firms to study, often constraining researchers to small sample sizes for their studies.

The number of entrepreneurship centers in universities has increased dramatically over the last two decades (Cooper, 2003). For scholars interested in gaining access to potentially rich data that can become the source for forming and subsequently examining significant empirical research questions, these centers and their ability to link scholars with data are important. Indeed, it may be that entrepreneurship centers can be a source of "important" rather than merely "available" data for researchers (Coviello & Jones, 2004). On the other hand, one of the key disadvantages associated with available data is that they typically are used to form convenience samples rather than samples that are most appropriate to test the researcher's theoretically based expectations.

Over 15 years ago, Low and MacMillan (1988) underscored level of analysis, time frame, and methodology type as important methodological issues that were affecting empirical entrepreneurship work. A recent issue of *Entrepreneurship Theory and Practice* was devoted to a retrospective examination of entrepreneurship research. Scholars publishing in that issue

essentially supported the concerns articulated by Low and MacMillan (1988) and identified additional concerns as well. Chandler and Lyon (2001), for example, focused on issues such as the need for multiple-source data sets, reliability and validity problems, theoretical model development, and longitudinal research. In the same issue, Davidsson & Wiklund (2001) published an in-depth examination of levels of analysis issues in entrepreneurship research.

In their investigation of strategic management research as published in the first volume of this series, Hitt et al. (2004) addressed data issues including sampling, statistical power, construct measurement, and analysis. In subsequent work, Boyd, Gove and Hitt (2005) completed an in-depth examination of construct measurement issues in strategic management research. In part, their results show that strategic management researchers do not consistently discuss reliability and validity issues, and empirical work in the strategy field often relies on single-indicator measures. We believe that certainly similar concerns (such as those addressed by Hitt et al. (2004) and Boyd et al. (2005)) apply to the entrepreneurship field. More particularly, however, our concentration is on four issues that we believe are critical to enhancing methodological rigor in entrepreneurship research: sampling, measurement, time frame, and effect size.

To compile data for our analysis, we reviewed all empirical papers published from 1999–2003 in *Journal of Business Venturing* and *Entrepreneurship Theory and Practice*. However, we examined only *JBV* empirical articles for 2004 and the first two issues of 2005 as the foundation for observations we have to offer about the reporting and discussion of effect size. We chose this sampling frame to focus on the most current work available to us. Our thought is that the earlier empirical work in our broader sample is less likely to have reported and certainly to have discussed effect size.

Initially, *Journal of Business Venturing* published only quantitative research, meaning that this dedicated journal is the one in which the research methods being used by entrepreneurship researchers were highly visible. However, *Entrepreneurship Theory & Practice* and certainly the journals covering broader management topics are additional outlets in which scholars can learn about the methods entrepreneurship researchers have used and are using to complete their work. Using ABI/Inform, we also identified papers published (during the same time period) in *Academy of Management Journal*, *Administrative Science Quarterly*, *Management Science*, *Organization Science*, and *Strategic Management Journal* that contained the word "entrepreneur" or one of its derivatives in either the paper's title or abstract for the same time frame.

While our sampling design may not have allowed us to identify every empirical paper published during the selected time period and in the chosen journals, we believe that it resulted in a representative sample. Articles that were primarily descriptive in nature, based solely on case study analysis, solely theoretical, or were based on experimental designs were excluded. This sampling procedure resulted in a sample of 158 articles. The numbers of empirical articles selected from the various journals were as follows: *Entrepreneurship Theory and Practice*–55; *Journal of Business Venturing*–79; *Academy of Management Journal*–5; *Administrative Science Quarterly*–2; *Management Science*–5; *Organization Science*–3; and *Strategic Management Journal*–9.

Sampling

Unlike strategic management research (Hitt et al., 2004), entrepreneurship research primarily involves the use of relatively small samples that are drawn principally from surveys (Chandler & Lyon, 2001). The sampling procedure used in a majority of entrepreneurship empirical work suggests concerns about the generalizability of findings. Moreover, some researchers' failure to carefully discuss or describe their sampling procedure is also a concern.

Grounded in established methods work, Hitt and colleagues (2004) identified two basic sampling designs – probability sampling (i.e., random sampling) and nonprobability sampling – that strategy researchers tend to use. The major benefit of random or stratified sampling is the ability to exclude systematic errors, while nonprobability sampling requires much more interpretation of both data and results (Hitt et al., 2004; Kerlinger & Lee, 2000). In our review of the entrepreneurship research included in our sample, 44% of the studies used either a random or stratified sample. This seems consistent with a recent content analysis of published papers on organizational performance determinants that reported a heavy reliance on purposive samples with a focus on available data (Short, Ketchen & Palmer, 2002). The use of purposive samples makes it difficult to compare and contrast results between and among studies examining particular relationships.

A related problem with sampling techniques is sample representativeness. Whether using a random, stratified or nonrandom sample, it is important for entrepreneurship researchers to attempt to establish that their data are representative of the universe from which they were drawn. This is particularly important due to the heavy reliance in entrepreneurship research on survey data. Issues such as nonrespondent bias and self-selection bias are important to evaluate with the intent of establishing acceptable levels of

validity. Among the papers reviewed here, approximately 30% of the studies explicitly reported either non-respondent or self-selection bias tests and statistics. It was common for researchers to report that they had carried out these tests and found no bias. However, it was also common for these researchers to omit reporting test descriptions and statistics to support their assertions. In some instances, researchers reported that their data had been validated in prior studies but again failed to offer statistics and results supporting those claims.

Effectively dealing with the related issues of sample criteria and sampling design can be challenging, especially in disciplines where paradigms are still evolving (Coviello & Jones, 2004). Nonetheless, excellent examples exist for entrepreneurship scholars to consider when dealing with sampling issues. For example, Barringer and Bluedorn (1999) developed a stratified random sample of manufacturing firms in the Midwestern and Southwestern regions of the United States. The authors developed, pre-tested, and mailed surveys to 501 firms using a modified Dillman (1978) method. Following the first survey mailing, a second, identical survey was mailed to another top manager in each of the responding firms. Results from the two mailings were assessed to establish inter-rater reliability. The researchers also used and reported results for Harman's one-factor test to check for common method bias and used three relevant characteristics to examine potential non-response bias problems. Based on these tests, the authors concluded that their data was representative and that it did not suffer from common method bias and non-response bias. Sampling procedures such as those used by Barringer and Bluedorn (1999) exemplify how entrepreneurship samples can be formed and assessed.

Measurement

Researchers in both strategic management (Boyd et al., 2005; Hitt et al., 2004; Hitt et al., 1998) and entrepreneurship (Chandler & Lyon, 2001) have noted a lack of attention to reliability and validity in construct measurement. We make the same observation here. Reliability and validity are important issues. When they are not carefully considered, reliability and validity issues create the need for significantly larger samples and can lead to insufficient statistical power and higher incidences of Type I and Type II errors (Hitt et al., 2004).

Consistent with results reported herein, Chandler and Lyon (2001) note that empirical entrepreneurship research relies heavily on survey data. This makes construct measurement particularly critical. Hitt et al. (2004) reported the same concern regarding strategic management research. Secondary data, while generally assumed to be reliable and accurate, can be subject to

the same bias associated with survey data. Because of this, researchers should engage in efforts to validate secondary data. Chandler and Lyon (2001) reported both instances of construct reliability and validity. They found that 40% of the studies using survey methods they reviewed used multiple item scales and provided coefficient alphas or some other internal consistency measure. Of the studies reported in the articles Chandler and Lyon (2001) surveyed, 29% reported alphas less than the widely accepted 0.70 (Nunnally, 1978) for more than 25% of their measures. While alpha coefficients are widely used to establish construct reliability, confirmatory factor analysis allows researchers to add theoretical arguments to their constructs and provides cross-validation that allows a test of the full model against an alternative model (Busenitz, Gomez & Spencer, 2000). Cross-validation procedures have a long history in both econometric and psychometric studies (Bagozzi & Yi, 1988) but are rare in entrepreneurship (Chandler & Lyon, 2001) or strategic management (Hitt et al., 2004) research. In our review of empirical entrepreneurship studies, we searched for papers that went beyond the use of alpha coefficients and made use of confirmatory factor analysis or other means to establish reliability such as using secondary data to validate survey data. In our set of 158 papers, only 27 (17%) used these or other additional means to establish reliability.

We can highlight the work of two sets of researchers to demonstrate how reliability should be established in entrepreneurship studies. In their work, Busenitz et al. (2000) introduced and validated a measure of country institutional profile for entrepreneurship. To establish reliability, they performed a double cross-validation procedure in which they split the sample and performed confirmatory factor analyses on each sample to ensure model validity. Zahra and Garvis (2000) surveyed 600 manufacturing firms and used measures for international corporate entrepreneurship (Miller, 1983) among others. They validated this measure with data from COMPUSTAT, *Fortune 500*, *Global Business 1000*, *Global Scope*, and *Forbes*. This attention to reliability is necessary to ensure that the constructs entrepreneurship researchers develop and use are accurate/appropriate and do not suffer from significant biases.

Construct validity is equally important. Validity ensures that researchers are actually measuring the intended construct (Chandler & Lyon, 2001). While construct validity is an essential aspect of empirical research, Chandler and Lyon (2001) reported that only 26% of the studies they reviewed paid attention to validation procedures beyond face validity. Issues of content validity, substantive validity, structural validity, and external validity are each important, yet are infrequently assessed in entrepreneurship research.

For example, only 31% of the papers reviewed by Chandler and Lyon (2001) showed that primary data through survey methodologies used factor analysis or reported the results of factor analysis done in prior studies while also providing evidence of convergent and discriminant validity. Busenitz et al. (2000) have provided an excellent example of how to establish construct validity. These authors discuss and test for external (by using several ANOVA tests), convergent (by using archival data on constructs the researchers expected to be logically related to dimensions of their country institutional profile), and predictive (by using two measures of entrepreneurship – a key construct) validity. This process enhances the confidence in the constructs they developed significantly and provides guidance to future researchers who desire to develop and use reliable measures and constructs to replicate or extend the Busenitz et al. (2000) results.

Time Frame

Low and MacMillan (1988), Davidsson, Low and Wright (2001), and Chandler and Lyon (2001) each note the need for a larger number of longitudinal designs to assess entrepreneurship questions. This view is consistent with the Shane and Venkataraman (2000) definition of entrepreneurship. Their definition explicitly states that entrepreneurship is concerned with, "how, by whom, and with what effects opportunities to create future goods and services are discovered, evaluated, and exploited" (Shane and Venkataraman, 2000, p. 218). This definition suggests that entrepreneurship is largely a process of discovery, evaluation, and opportunity exploitation that occurs over time. Low and MacMillan (1988) suggest that understanding this pattern of development necessitates the use of longitudinal designs even though longitudinal designs in entrepreneurship pose a number of difficult problems, including the demands of tenure (Davidsson et al., 2001), the lack of available time-series data (Davidsson et al., 2001), high failure rates (Davidsson et al., 2001; Low & MacMillan, 1988), and high costs (Low & MacMillan, 1988). These difficulties appear evident as only 7% of the studies examined by Chandler and Lyon (2001) were longitudinal studies where data were collected over two or more points in time. Our own review of the literature shows that 13% of the studies conducted in the last 5 years were longitudinal, suggesting that we as entrepreneurship researchers are making slow progress toward one of the Low and MacMillan (1988) major recommendations.

Effect Sizes

Reporting effect sizes in a study in lieu of, or at least in addition to null hypothesis significance testing (NHST) may facilitate the linkage between

empirical work and theoretical frameworks that entrepreneurship researchers can use to specify testable relationships. Effect sizes, which have been categorized into over 3 dozen types (Kirk, 1996), quantify the degree to which sample results diverge from the null hypothesis (Cohen, 1994; Thompson, 2005). The American Psychological Association's (2001, p. 599) Task Force on Statistical Inference suggested that "reporting and interpreting effect sizes in the context of previously reported effects is essential to good research." The debate regarding the value created by reporting effects sizes versus reporting the results of testing null hypotheses is long lived (Cohen, 1994), and empirical studies highlighting the reporting of NHST without effect sizes have grown exponentially and across a wide range of disciplines (Anderson, Burnham & Thompson, 2000), from economics (e.g., Ziliak & McCloskey, 2004), education (e.g., Thompson, 1996), psychology (e.g., Schmidt, 1996), to the wildlife sciences (e.g., Anderson et al., 2000).

These suggestions are interesting within the context of our sample of empirical work published in *JBV* for 2004 and the first two issues of 2005. While virtually all of the studies reported effect size, a very small percentage discussed the implications of their analysis to assess effect size. Moreover, very few studies compared their effect-size results to previous work. Thus, an admittedly limited examination of recent entrepreneurship publications suggests that researchers are reporting effect size. This is a positive outcome in light of the Anderson et al. (2000) observation. However, entrepreneurship researchers should take the next steps of fully discussing the implication of the results of their analysis of effect size within the context of previously reported effects.

A number of forces are driving the shift to the increased reporting of effect sizes. Some (e.g., Cohen, 1994) argue that widespread confusion and misinterpretation underlie the concept of statistical significance and conclusions regarding decisions about the null hypothesis. Addressing the issue of statistical significance, the actual implications of rejecting the null hypothesis at a given p value is that there is a probability p for the given sample statistics at the measured sample size assuming the null hypothesis is true in the population. However, because the direction of inference is from the population (and the sampling distribution) to the sample and not vice versa, making assertions regarding the population or generalizing across other samples is purely speculative and groundless when based solely on a value of p. Furthermore, p is not the probability that the results are by chance, nor is $(1-p)$ the probability that the result will be replicated under constant conditions (Kline, 2004).

A final point should be made regarding statistical significance. To some extent, minimal as it might be, all null hypotheses are false. As Cohen (1994, p. 1000) argues, "If it is false, even to a tiny degree, it must be the case that a large enough sample will produce a significant result and lead to its rejection. So if the null hypothesis is always false, what's the big deal about rejecting it?"

Further issues arise with the meaning of a particular decision regarding the null hypothesis. For example, one common misconception is that a rejection of the null hypothesis infers that the underlying causal mechanism is identified (Kline, 2004). Additional misinterpretations about null hypothesis decisions are: (1) that there is no population effect size if the study fails to reject the null, (2) that the alternative hypothesis is confirmed with the rejection of the null, and (3) that the study is a failure if the null hypothesis is not rejected (Kline, 2004). As noted previously, rejecting the null hypothesis (i.e., having statistical significance) signifies the probability for the given sample statistics for the sample size assuming that the null hypothesis is true in the population. This statement in no way allows one to make assertions regarding the population or any other sample.

Some argue that even understanding statistical significance does not tell us what we really want to know. As Thompson (2005) notes, "If a literature involved 100 studies of an important outcome, each yielding identical, nonzero $\underline{d}$ values (effect sizes), would a scholar care that in all studies p equaled 0.06? Do we ultimately care about p or do we care about effects and their replicability?" Invariably, our interest lies in the effect size – the extent to which the independent variable impacts the dependent variable (Kline, 2004). Strong theory articulates not only what causal relationships exist but also the magnitude of these relationships. The dichotomous approach (i.e., reject or fail to reject the null hypothesis) of NHST allows one to only speculate about the relationships in the overall population. However, this conjecture is not grounded in the statistical process, nor does NHST provide any information on what the magnitude of the conjecture is. In other words, NHST fails to support even weak theory.

Reporting and discussing effect sizes is particularly relevant to building strong theory as the foundation for significant empirical entrepreneurship research. Effect sizes not only highlight specific relationships but also quantify the magnitude of the relationships. However, the distinctive value of reporting effect sizes is actually due to the available sample size characteristic of entrepreneurship research. As noted earlier, empirical entrepreneurship studies often must accommodate dynamic markets and technologies, fleeting opportunities, high rates of organizational failure, and the time and

willingness of entrepreneurs as participants in the researcher's study. These factors severely limit the sample size that is available to the researcher. While this greatly impedes one's ability to achieve statistical significance, effect sizes are not solely a function of sample size. Rosnow and Rosenthal (1989, p. 1279) point out that, "… it is important to realize that the effect size tells us something very different from the *p* level. A result that is statistically significant is not necessarily practically significant as judged by the magnitude of the effect. Consequently, highly significant *p* values should not be interpreted as automatically reflecting large effects." In other words, while a study with a small sample size may have a small calculated *p* value, the measured effect size may actually be quite large – and vice versa. Because of this, entrepreneurship theory can be at least partly built on empirical studies that are constrained to small sample sizes but report and theoretically discuss measured effect sizes.

A number of points should be mentioned regarding the use of effect sizes when engaging in empirical entrepreneurship research. First, effect size can be calculated multiple ways. Therefore, specification of the particular effect size measured should be explicitly stated so that comparisons among effect sizes across studies are appropriate. A second suggestion is that effect sizes should be reported alongside results for statistical significance and results that are not statistically significant. As observed above by Thompson (2005) and Rosnow and Rosenthal (1989), results that are not statistically significant can still have what is considered a large effect size. Neither statistical significance nor the lack thereof is synonymous with the importance of results. Neglecting to report effect sizes corresponding to results that are not significant, therefore, is discarding valuable information. Finally, ambiguity continues to surround the debate on what is a worthy effect size. What constitutes a relatively considerable effect size will depend largely on what is being studied. In some studies, the independent variable may be expected to have a small effect on the dependent variable, so a large effect would be seemingly alarming. As in reporting any form of statistics, it is important to understand the implications made and how the findings fit with expectations. The theoretical development of empirical studies, therefore, should include an articulation of the reasoning behind one's expectations for the magnitude of a measured effect.

For those interested in studying this issue in greater detail, work by Kirk (2003), Kline (2004), Snyder and Lawson (1993) and Thompson (2005) can be consulted. In total, these researchers' discussions and analyses provide us with accessible, useful explanations of effect sizes and their use. We present a summary of the methodological issues discussed herein in Table 2.

Table 2. Methodological Challenges and Recommendations for Entrepreneurship Research.

	Challenges and Recommendations
Sampling	Greater use of random or stratified samples Testing for sample representativeness Inadequate sample sizes to defect Type I and Type II errors
Measurement	Reliability and validity must be established to minimize the potential for Type I and Type II errors When possible, use multiple indicators for constructs Cross-validation with secondary data or by splitting the sample enhances validity Cronbach's alpha alone is not sufficient
Time frame	Time-series data are needed to establish causality Entrepreneurship is a process that develops over time, suggesting the need for properly designed longitudinal work
Effect size	Report and discuss effects sizes When possible, compare effect sizes across studies

FINAL RECOMMENDATIONS

We agree with Shane and Venkataraman (2000, p. 224) who argue that "entrepreneurship is an important and relevant field of study." Our purpose in evaluating entrepreneurship research is not to highlight the shortcomings of previous research. Rather, our interest is to pinpoint some theoretical and methodological issues that may be stifling scholars' efforts to develop a systematic body of information about entrepreneurship. Shane and Venkataraman (2000) note that skeptics claim it is impossible to develop a systematic body of entrepreneurship knowledge that can be used to inform the specification of research in this domain. In spite of such criticism, we believe entrepreneurship researchers can prove the skeptics wrong through a greater emphasis on theoretical and empirical rigor. A final set of recommendations is presented below, which seeks to highlight the actions researchers can take to develop a systematic body of knowledge and expand it in a cumulative manner. We hope that these recommendations, many of which are based on the pioneering work of a host of scholars, will stimulate the development of interesting questions for researchers to specify and then examine.

1. *Articulate* "*how*," "*why*," *and* "*when*." Research should identify relationships between established theory and assertions being made and

articulate the reasoning underlying the relationships on the bases of how, why, and when. A systematic body of knowledge implies an understanding of relationships among a wide range of variables. Individual theoretical and empirical works are often not afforded the space to develop a full body of knowledge. Therefore, it is important for scholars to relate their focused assertions to an established stream of knowledge.

2. *Seek to develop theoretical frameworks specific to entrepreneurship.* Although recommendation 1 above argues for a more incremental approach to theory development, entrepreneurship remains a young field of research that is characterized by an incompletely specified paradigm. Shane and Venkataraman (2000) recommended the development of separate theories of the entrepreneur and of opportunity. While a theory of the entrepreneur may prove fruitful, one could also argue that personality traits are stable; therefore, one could argue that an individual is or is not entrepreneurial, and changes to this characterization are unlikely. In this case, theory development surrounding the *context* within which entrepreneurs surface may be more productive. For example, greater interest would lie in *how* an entrepreneur can enhance his or her entrepreneurial capacity. In addition, the actions institutional actors (e.g., universities, incubators, economic development groups) take to support and sustain entrepreneurial behaviors are especially informative. A second opportunity for research is to form a theory of corporate entrepreneurship that could then be empirically assessed. Many of the factors, such as incentive systems and the lack of autonomy, present in large corporations oppose entrepreneurial behaviors. The development of a theory of corporate entrepreneurship could facilitate our understanding of how corporations can manage their resources to promote entrepreneurial actions. In other words, can a group of non-entrepreneurs be organized in a way that their collective behaviors are entrepreneurial while their individual behaviors are not? These paths offer different perspectives on entrepreneurship. But while developing new theories may be likened to radical innovations, the relationships among variables must be explicated along the lines of how, why, and when.
3. *Report and discuss effect sizes.* It is important to report effect sizes for all empirical findings, including those that are statistically insignificant. Effect sizes quantify the magnitude of relationships, thereby facilitating the development of strong theory. Nevertheless, the magnitude of the effect size should be discussed relative to the scholar's expectations and, when possible, compared with previously reported effect sizes for a given relationship.
4. *Enhance procedural research activities.* Procedural research activities include referencing, illustrating arguments, and presenting quantitative

information in a comprehensible manner. Harzing (2002), for example, describes appropriate referencing techniques as well as the consequences of inaccurate referencing. A legitimate systematic body of knowledge grows from the ability to compare and contrast theory and empirical results within the context of previously reported findings. Procedural research activities facilitate scholars' efforts to effectively present their carefully constructed theoretical arguments as well as to interpret their empirical findings in ways that can inform future research.

CONCLUSION

Increasingly, entrepreneurship and entrepreneurial behavior are being recognized as the source of growth and development for established and emerging economies throughout the world. One reason for this recognition is that the often positive effect of entrepreneurship on the productivity of global economies is being widely established in a variety of contexts. Nevertheless, the high failure rate of entrepreneurial ventures signifies the importance for scholars to develop theoretically rich research questions that can be examined through rigorous and appropriate empirical designs. As a guide, entrepreneurial research must become theory-centric, driven by the need for a cumulative nature to research that develops along common lines of thought. Such theoretical lenses position the field to better articulate entrepreneurial phenomena and guide both future research and provide insights that have the potential to positively contribute to the practice of entrepreneurship. On the other hand, theory-driven research alone is susceptible to criticism. To serve as a *valid* guide to entrepreneurial efforts and future research, entrepreneurial theories must be examined or tested by using appropriate methodologies. Theory and methodology are symbiotic components of research, with theory serving as a common foundation for the specification of empirical research with the potential to yield valid, reliable results on which future studies can be built. Employing the appropriate methods allows researchers to ascertain to a greater degree the validity of empirical results and more accurately translate findings into theoretical extensions and evaluations.

REFERENCES

Adler, P. S., & Kwon, S.-W. (2002). Social capital: Prospects for a new concept. *Academy of Management Review*, *27*, 17–40.

Alvarez, S. A., & Barney, J. B. (2001). How entrepreneurial firms can benefit from alliances with large partners. *Academy of Management Executive*, *15*, 139–148.

Alvarez, S. A., & Barney, J. B. (2004). Organizing rent generation and appropriation: Toward a theory of the entrepreneurial firm. *Journal of Business Venturing*, *19*, 621–635.

American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.

Amit, R., Glosten, L., & Muller, E. (1993). Challenges to theory development in entrepreneurship research. *Journal of Management Studies*, *30*, 815–834.

Amit, R., & Schoemaker, P. J. H. (1993). Strategic assets and organizational rent. *Strategic Management Journal*, *14*, 33–46.

Anderson, D. R., Burnham, K. P., & Thompson, W. (2000). Null hypothesis testing: Problems, prevalence, and an alternative. *Journal of Wildlife Management*, *64*, 912–923.

Bacharach, S. B. (1989). Organizational theories: Some criteria for evaluation. *Academy of Management Review*, *14*, 496–515.

Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, *16*(1), 74–94.

Barney, J. B. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, *17*, 99–120.

Barringer, B. R., & Bluedorn, A. C. (1999). The relationship between corporate entrepreneurship and strategic management. *Strategic Management Journal*, *20*, 421–444.

Batjargal, B., & Liu, M. (2004). Entrepreneurs' access to private equity in China: The role of social capital. *Organization Science*, *15*, 159–172.

Bergh, D. D. (2003). Thinking strategically about contribution. *Academy of Management Journal*, *46*, 135–136.

Blalock, H. M. (1979). *Social statistics* (2nd ed.). New York: McGraw-Hill.

Boyd, B. P., Gove, S., & Hitt, M. A. (2005). Construct measurement in strategic management research: Illusion or reality? *Strategic Management Journal*, *26*, 239–257.

Burt, R. S. (1992). *Structural holes: The social structure or competition*. Cambridge, MA: Harvard University Press.

Busenitz, L. W., Gomez, C., & Spencer, J. W. (2000). Country institutional profiles: Unlocking entrepreneurial phenomena. *Academy of Management Journal*, *43*, 994–1003.

Busenitz, L. W., West, G. P., III, Shepherd, D., Nelson, T., Chandler, G. N., & Zacharakis, A. (2003). Entrepreneurship research in emergence: Past trends and future directions. *Journal of Management*, *29*, 285–308.

Chandler, G. N., & Lyon, D. W. (2001). Issues of research design and construct measurement in entrepreneurship research: The past decade. *Entrepreneurship Theory and Practice*, *25*(4), 101–113.

Chrisman, J. J., & Kellermanns, F. (2005). Entrepreneur. In: M. A. Hitt & R. D. Ireland (Eds), *The Blackwell Encyclopedia of Management, Entrepreneurship*, (2nd ed.) (pp. 61–63). Malden, MA: Blackwell Publishing.

Cohen, J. (1994). The earth is round ($p<0.05$). *American Psychologist*, *49*, 997–1003.

Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, *94*(Supplement), S95–S120.

Combs, J. G., Ketchen, D. J., Jr., & Hoover, V. L. (2004). A strategic groups approach to the franchising-performance relationship. *Journal of Business Venturing*, *19*, 877–897.

Cooper, A. (2003). Entrepreneurship: The past, the present, the future. In: Z. J. Acs & D. B. Audretsch (Eds), *Handbook of entrepreneurship research* (pp. 21–34). Boston: Kluwer Academic Publishers.

Coviello, N. E., & Jones, M. V. (2004). Methodological issues in international entrepreneurship research. *Journal of Business Venturing*, *19*, 485–508.

Davidsson, P., Low, M. B., & Wright, M. (2001). Editor's introduction: Low and MacMillan ten years on: Achievements and future directions for entrepreneurship research. *Entrepreneurship Theory and Practice*, *25*(4), 5–15.

Davidsson, P., & Wiklund, J. (2001). Levels of analysis in entrepreneurship research: Current research practice and suggestions for future research. *Entrepreneurship Theory and Practice*, *25*(4), 81–99.

Dew, N., Velamuri, S. R., & Venkataraman, S. (2004). Dispersed knowledge and an entrepreneurial theory of the firm. *Journal of Business Venturing*, *19*, 659–679.

Dierickx, I., & Cool, K. (1989). Asset stock accumulation and sustainability of competitive advantage. *Management Science*, *35*, 1504–1513.

Dillman, D. A. (1978). *Main and telephone surveys: The total design method*. New York: Wiley.

DiMaggio, P. J. (1995). Comments on "What theory is not". *Administrative Science Quarterly*, *40*, 391–397.

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, *48*, 147–160.

Dubin, R. (1969). *Theory Building*. New York: Free Press.

Eckhardt, J. T., & Shane, S. A. (2003). Opportunities and entrepreneurship. *Journal of Management*, *29*, 333–349.

Gartner, W. B. (1990). What are we talking about when we talk about entrepreneurship? *Journal of Business Venturing*, *5*, 15–28.

Gartner, W. B. (2001). Is there an elephant in entrepreneurship? Blind assumptions in theory development. *Entrepreneurship Theory and Practice*, *25*(4), 27–39.

Grant, R. M. (1996). Knowledge, strategy, and the theory of the firm. *Strategic Management Journal*, *17*, 109–122.

Grimm, C. M., & Smith, K. G. (1997). *Strategy as action: Industry rivalry and coordination*. Cincinnati, OH: SouthWestern College Publishing Company.

Guler, I., Guillen, M. F., & Macpherson, J. M. (2002). Global competition, institutions, and the diffusion of organizational practices: The international spread of ISO 9000 quality certificates. *Administrative Science Quarterly*, *47*, 207–232.

Harrison, R. T., & Leitch, C. M. (1996). Discipline emergence in entrepreneurship: Accumulative fragmentalism or paradigmatic science? *Entrepreneurship, Innovation, and Change*, *5*(2), 65–83.

Hart, S., & Banbury, C. (1994). How strategy-making processes can make a difference. *Strategic Management Journal*, *15*, 251–269.

Hart, S. L. (1992). An integrative framework for strategy-making processes. *Academy of Management Review*, *17*, 327–351.

Harzing, A.-W. (2002). Are our referencing errors undermining our scholarship and credibility? The case of expatriate failure rates. *Journal of Organizational Behavior*, *23*, 127–148.

Hitt, M. A., Boyd, B. K., & Li, D. (2004). The state of strategic management research and a vision of the future. In: D. J. Ketchen, Jr., & D. D. Bergh (Eds), *Research methodology in strategy and management* (Vol. 1, p. 1–31).

Hitt, M. A., Gimeno, J., & Hoskisson, R. E. (1998). Current and future research methods in Strategic management. *Organizational Research Methods*, *1*, 6–44.

Hrebiniak, L. G., & Joyce, W. F. (2001). Implementing strategy: An appraisal and agenda for future research. In: M. A. Hitt, R. E. Freeman & J. S. Harrison (Eds), *The Blackwell Handbook of strategic management* (pp. 602–626). Malden, MA: Blackwell Business.

Ireland, R. D., Hitt, M. A., & Sirmon, D. G. (2003). A model of strategic entrepreneurship: The construct and its dimensions. *Journal of Management*, *29*, 1–26.

Ireland, R. D., Hitt, M. A., & Webb, J. W. (2005). Entrepreneurial alliances and networks. In: J. J. Reuer & O. Shenkar (Eds), *Handbook of Strategic Alliances*. Sage Publishers (in press).

Jones, G. R. (2001). Towards a positive interpretation of transaction cost theory: The central roles of entrepreneurship and trust. In: M. A. Hitt, R. E. Freeman & J. S. Harrison (Eds), *The Blackwell handbook of strategic management* (pp. 208–228). Malden, MA: Blackwell Business.

Kerlinger, F. M., & Lee, H. B. (2000). *Foundations of behavioral research* (4th ed.). Belmont, CA: Wadsworth/Thomson Learning.

Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, *56*, 746–759.

Kirk, R. E. (2003). The importance of effect magnitude. In: S. F. Davis (Ed.), *Handbook of research methods in experimental psychology* (pp. 83–105). Oxford, United Kingdom: Blackwell.

Kirzner, I. M. (1979). *Perception, opportunity and profit*. Chicago: University of Chicago Press.

Kline, R. B. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research*. Washington, DC: American Psychological Association.

Koestler, A. (1964). *The act of creation*. New York: Dell.

Kogut, B., & Zander, U. (1992). Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science*, *3*, 383–397.

Kondra, A. Z., & Hinings, C. R. (1998). Organizational diversity and change in institutional theory. *Organization Studies*, *19*, 743–767.

Kuhn, T. S. (1996) (3rd ed.). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Low, M. B., & MacMillan, I. C. (1988). Entrepreneurship: Past research and future challenges. *Journal of Management*, *14*, 139–161.

Martinelli, A. (1994). Entrepreneurship and management. In: N. Smelser & R. Swedberg (Eds), *Handbook of Economic Sociology* (pp. 476–503). Princeton, NJ: Princeton University Press.

Miller, D. (1983). The correlates of entrepreneurship in three types of firms. *Management Science*, *29*, 770–791.

Mintzberg, H. (1973). Strategy-making in three modes. *California Management Review*, *16*, 44–53.

Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.

Priem, R. L., & Butler, J. E. (2001a). Is the resource-based "view" a useful perspective for strategic management research? *Academy of Management Review*, *26*, 22–40.

Priem, R. L., & Butler, J. E. (2001b). Tautology in the resource-based view and the implications of externally determined resource value: Further comments. *Academy of Management Review*, *26*, 57–66.

Reed, R., & DeFillippi, R. J. (1990). Causal ambiguity, barriers to imitation, and sustainable competitive advantage. *Academy of Management Review*, *15*, 88–102.

Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist*, *44*, 1276–1284.

Rynes, S. (2002). Some reflections on contribution. *Academy of Management Journal*, *45*, 311–313.

Schmidt, F. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for the training of researchers. *Psychological Methods*, *1*, 115–129.

Schumpeter, J. A. (1934). *The theory of economic development*. Cambridge, MA: Harvard University Press.

Sexton, D. L. (1988). The field of entrepreneurship: Is it growing or just getting bigger? *Journal of Small Business Management*, *26*, 5–8.

Shane, S. (2000). Prior knowledge and the discovery of entrepreneurial opportunities. *Organization Science*, *11*, 448–469.

Shane, S., & Stuart, T. (2002). Organizational endowments and the performance of university start-ups. *Management Science*, *48*, 154–170.

Shane, S., & Venkataraman, S. (2000). The promise of entrepreneurship as a field of research. *Academy of Management Review*, *25*, 217–226.

Sharma, P., & Chrisman, J. J. (1999). Toward a reconciliation of the definitional issues in the field of corporate entrepreneurship. *Entrepreneurship Theory and Practice*, *23*(3), 11–27.

Short, J. C., Ketchen, D. J., & Palmer, P. B. (2002). The role of sampling in strategic management research on performance: A two-study analysis. *Journal of Management*, *28*, 363–385.

Sirmon, D. G., Hitt, M. A., & Ireland, R. D. (2005). Managing firm resources in dynamic environments to create value: Looking inside the black box. *Academy of Management Review* (in press).

Slack, T., & Hinings, B. (1994). Institutional pressures and isomorphic change: An empirical test. *Organization Studies*, *15*, 803–827.

Smith, K. G., & Di Gregorio, D. (2002). Bisociation, discovery, and the role of entrepreneurial action. In: M. A. Hitt, R. D. Ireland, S. M. Camp & D. L. Sexton (Eds), *Strategic entrepreneurship: Creating a new mindset* (pp. 129–150). Oxford: Blackwell Publishers.

Snyder, P., & Lawson, S. (1993). Evaluating results using corrected and uncorrected effect size estimates. *Journal of Experimental Education*, *61*, 334–349.

Steier, L., & Greenwood, R. (2000). Entrepreneurship and the evolution of angel financial networks. *Organization Studies*, *21*, 163–192.

Stevenson, H. H., & Jarillo, J. C. (1990). A paradigm of entrepreneurship: Entrepreneurial management. *Strategic Management Journal*, *11*, 17–27.

Sutherland, J. W. (1975). *Systems: Analysis, Administration, and Architecture*. New York: Van Norstrand.

Sutton, R. I., & Staw, B. M. (1995). What theory is not. *Administrative Science Quarterly*, *40*, 371–384.

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, *18*, 509–533.

Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms. *Educational Researcher*, *25*(2), 26–30.

Thompson, B. (2005). Research synthesis: Effect sizes. In: J. Green, G. Camilli & P. B. Elmore (Eds), *Complementary methods for research in education*. Washington, DC: American Educational Research Association, Forthcoming.

Uzzi, B. (1997). Social structure and competition in interfirm networks: The paradox of embeddedness. *Administrative Science Quarterly*, *42*, 35–67.

Venkataraman, S. (1997). The distinctive domain of entrepreneurship research. In: J. Katz & R. Brockhaus (Eds), *Advances in entrepreneurship, firm emergence and growth*, (Vol. 3, pp. 119–138). Greenwich: JAI Press.

Venkataraman, S., & Harting, T. (2005). Entrepreneurial opportunity. In: M. A. Hitt & R. D. Ireland (Eds), *The Blackwell Encyclopedia of Management, Entrepreneurship*, (2nd ed.) (pp. 100–103). Malden, MA: Blackwell Publishing.

Wacker, J. G. (2004). A theory of formal conceptual definitions: Developing theory-building measurement instruments. *Journal of Operations Management*, *22*, 629–650.

Webb, J. W., & Ireland, R. D. (2004). Resource efficiency of start-up ventures: The link to effectiveness in the terms of creative destruction process. Paper presented at the babson kauffman entrepreneurship research conference.

Weick, K. E. (1989). Theory construction as disciplined imagination. *Academy of Management Review*, *14*, 516–531.

Weick, K. E. (1995). What theory is not, theorizing is. *Administrative Science Quarterly*, *40*, 385–390.

Weick, K. E. (1999). Theory construction as disciplined reflexivity: Tradeoffs in the 90s. *Academy of Management Review*, *24*, 797–806.

Wernerfelt, B. (1984). A resource-based view of the firm. *Strategic Management Journal*, *5*, 171–180.

Whetten, D. A. (1989). What constitutes a theoretical contribution? *Academy of Management Review*, *14*, 490–495.

Zaheer, A., & Zaheer, S. (1997). Catching the wave: Alertness, responsiveness, and market influence in global electronic networks. *Management Science*, *43*, 1493–1509.

Zahra, S. A., & Garvis, D. M. (2000). International corporate entrepreneurship and firm performance: The moderating effect of international environmental hostility. *Journal of Business Venturing*, *15*, 469–492.

Zahra, S. A., Jennings, D. F., & Kuratko, D. F. (1999). The antecedents and consequences of firm-level entrepreneurship: The state of the field. *Entrepreneurship Theory & Practice*, *23*(3), 45–65.

Ziliak, S. T., & McCloskey, D. N. (2004). Size matters: the standard error of regressions in the *American Economic Review*. *Journal of Socio-Economics*, *33*, 527–546.

THE PROBLEM OF METHOD AND THE PRACTICE OF MANAGEMENT RESEARCH

Kent D. Miller

ABSTRACT

This chapter highlights the personal side of research methods. We begin with an overview of Hans-Georg Gadamer's insights into the general problem of method in the social sciences and hermeneutics. This is followed by an overview of Michael Polanyi's explanation of the practice of scientific investigation. The second half of the chapter considers implications of the personal side of methods for how we conduct management research. This section discusses critical realism as a philosophy of science consistent with the assumptions of our field, the reasons for methodological pluralism and possible responses, and management research as a social practice.

We usually leave ourselves out of discussions of research methods. Rarely do people show up in methods textbooks. We write the methods sections of our research articles using the third person, rather than offering first-person narratives. There are few biographical or autobiographical accounts of management researchers' experiences with methods. Those that we have

Research Methodology in Strategy and Management
Research Methodology in Strategy and Management, Volume 2, 143–177

ISSN: 1479-8387/doi:10.1016/S1479-8387(05)02007-2

(e.g. Simon, 1991; Starbuck, 2004) reveal the influences of mentors and colleagues, and personal judgments on methods choices, implementation, and interpretation of results. Barley (1990, p. 220) described his revealing account of personal experience with longitudinal field research methods in this way: "The telling weaves together the rational and the irrational, the planned and the unplanned, the personal and the impersonal to approach a sense of realism often lacking in the methods sections of most journal articles." Because such confessions violate norms of objectivity, most of us avoid drawing attention to our personal influence on the research process. By contrast, this chapter focuses on the personal and interpersonal aspects of research methods and their implications for how we do research.

To explore the human side of research methods, I draw upon the writings of philosopher Hans-Georg Gadamer on the general problem of method in the social sciences and hermeneutics and scientist-turned-philosopher Michael Polanyi on the practice of scientific research. Gadamer and Polanyi were influential twentieth-century thinkers who grappled with the implications of our human capacities and limitations for how we do research. Gadamer's work is probably unfamiliar to most management researchers, yet I intend to show that his discussion of methods in the social sciences and hermeneutics is quite relevant to our field. Among management researchers, Polanyi is associated with the concept "tacit knowledge," which figures prominently in discussions of organizational learning and routines, but few management scholars have explored the broader scope of his philosophical writings (Brown & Duguid, 2001; Foss, 2003; Tsoukas, 2003). Here we focus on his observations about the nature of scientific inquiry. What emerges from the writings of Gadamer and Polanyi is a portrayal of research methods as tools that we employ subjectively within local research traditions. Gadamer and Polanyi elaborated the roles of personal judgment and tacit knowledge in the practice of research. Both rejected objectivity as an unobtainable ideal, but they did so without falling into relativism, which would undermine science.

If our methods are unavoidably personal, what are the implications for management research? The latter half of this chapter discusses three broad implications. The first is a philosophy of science that recognizes that personal judgments are unavoidable. Although Gadamer and Polanyi preceded the seminal work on critical realism (see Archer, Bhaskar, Collier, Lawson, & Norrie, 1998), themes within their writings correspond closely with this philosophy of science. Critical realism acknowledges the subjectivity of scientists and holds to a realist ontology that makes science possible. I argue that most management research is compatible with critical realism.

A second area of implications is explaining the reasons for methodological pluralism within management research (see Hitt, Gimeno, & Hoskisson, 1998) and possible responses. Although management researchers openly acknowledge theoretical pluralism (e.g. Mintzberg, 1990) and its implications have been the subject of some debate (e.g. Canella & Paetzold, 1994; Montgomery, Wernerfelt, & Balakrishnan, 1989; 1991; Pfeffer, 1993; Seth & Zinkhan, 1991), the reasons for methodological pluralism and possible responses have not been discussed openly. To address methodological pluralism, management researchers may find the literature on hermeneutics helpful.

A third area of implications addresses management research as a social practice. If research processes cannot be reduced to codified methods, how do research practices emerge and evolve? What kinds of practices and social relationships are consistent with advancing knowledge? Polanyi's writings, along with the literature on communities of practice (e.g. Brown & Duguid, 1991, 2001) and epistemic cultures (Knorr Cetina, 1999) help us address these questions. Examining the practices of science reveals the integral connection between ethical values and advances in scientific knowledge.

The concluding section discusses specific applications for management researchers.

GADAMER ON THE PROBLEM OF METHOD

Who was Gadamer? Hans-Georg Gadamer (1900–2002) is recognized as one of the leading philosophers of the twentieth century. His primary contributions were to the field of hermeneutics.[1] In a recently published biography of Gadamer, Grondin described his magnum opus, *Truth and Method*, published in 1960, as "...the most important accomplishment in German philosophy since Heidegger's *Being and Time*" (2003, p. 291).

Going more directly to the interests of the readers of this volume, why does Gadamer's thought merit attention from management researchers?[2] In particular, what does he have to say that is relevant to research methods within our field? To both questions, I offer a preliminary response that many points raised by Gadamer bear upon what we do as management researchers and how we approach methods. To provide specific answers to these questions, we must move into the details of Gadamer's writings. Gadamer challenges us to recognize the implications of our human limitations for research methods. He offered compelling arguments that objectivity is an unattainable ideal for social scientists, and advocated that we acknowledge

the prejudgments that we bring to our research and our fallibility as interpreters of social phenomena. Gadamer's arguments also have implications for how we interact with one another within a research community.

Gadamer (2002) laid out the general problem of method in the social sciences (Geisteswissenschaften) in Part I of *Truth and Method*.[3] Gadamer used the term "method" to refer to a predetermined systematic investigative approach such as that applied in the natural sciences.[4] For Gadamer, the method of the natural sciences is inappropriate for the social sciences because assuming a posture of objectivity is inconsistent with understanding social phenomena. Social science requires sensitivity to the particulars of situations, rather than reliance upon general procedural rules and theoretical laws. The social science researcher exercises judgment that "...cannot be taught in the abstract but only practiced from case to case, and is therefore more an ability like the senses" (p. 31).[5] Rather than operating on the basis of objectivity, social scientists draw upon their own familiarity with social life acquired through personal experience. For Gadamer, this distinguishing characteristic of the social sciences, relative to the natural science, posed a problem of method.[6] If the method of the natural sciences is not a good fit with the character of social phenomena, how should we proceed with social science?

According to Gadamer, interest among social scientists in seeking to emulate the natural sciences by taking a detached objectivist posture can be traced to the influence of Immanuel Kant. Kant sought to establish transcendental principles for knowledge, ethics, and even esthetic judgments. On these subjects, Kant strove for universality, which required detachment from the idiosyncratic beliefs of particular communities. However, Kant recognized a residual category of judgments – tastes – that relied upon the preferences of local communities. Tastes cannot be specified by universal rules, but involve adopting a particular point of view (p. 32). Gadamer summarized:

> If we now examine the importance of Kant's *Critique of Judgment* for the history of the human sciences, we must say that his giving aesthetics a transcendental philosophical basis had major consequences and constituted a turning point. It was the end of a tradition but also the beginning of a new development. It restricted the idea of taste to an area in which, as a special principle of judgment, it could claim independent validity – and, by so doing, limited the concept of knowledge to the theoretical and practical use of reason. The limited phenomenon of judgment, restricted to the beautiful (and sublime), was sufficient for his transcendental purpose; but it shifted the more general concept of the experience of taste, and the activity of aesthetic judgment in law and morality, out of the center of philosophy (pp. 40–41).

Kant's critique of taste denied it any significance as knowledge (p. 43). Esthetic tastes can be shared, but they cannot be true in any sense that could be called scientific.

To illustrate the implications of the Kantian perspective, Gadamer turned to the question of artistic truth. Because the experience of art depends upon personal tastes, it cannot convey universal truth. Rather, its meaning can only be local and subjective. Art deals with appearances, not with reality (p. 82). Kant's perspective opened up an epistemic breach that privileged those categories of knowing that could appeal to universals relative to those that could only appeal to subjective local tastes. Only the former rises above one's own expectations and preferences.

Such a perspective sets up scientific inquiry as an idealized form of perception that is independent of tastes. Gadamer refuted such a view of science by denying that there are pure perceptions: "...our perception is never a simple reflection of what is given to the senses" (p. 90). Our understanding depends upon the preconceptions that we bring to the experience of perceiving. "Perception always includes meaning" (p. 92). As such, our perceptions involve an unavoidable interpretive (hermeneutical) dimension.

In Part II of *Truth and Method*, Gadamer developed the major points of his theory of hermeneutic experience. Gadamer's key move was to link hermeneutics to ontology.[7] Gadamer followed Heidegger (1962) in emphasizing the reality of the objects of study (i.e. the things themselves), which guide the interpreter: "For it is necessary to keep one's gaze fixed on the thing throughout all the constant distractions that originate in the interpreter himself" (p. 267). Only by attending to "the things themselves" can one's preconceptions be updated with new conceptions more consistent with the phenomenon of interest. Such an approach requires that we remain open to being surprised.

The fitting response to experience is to become teachable: "The dialectic of experience has its proper fulfillment not in definitive knowledge but in the openness to experience that is made possible by experience itself" (p. 355). We learn from experience through the disappointments (disconfirmed expectations) that it produces (p. 356). From experience, we come to know our own finiteness. The link back to ontology is essential: "Experience teaches us to acknowledge the real. The genuine result of experience, then – as of all desire to know – is to know what is" (p. 357).

Gadamer held that we always bring preconceptions (prejudgments) to our research.[8] This acknowledgment contradicts the scientific ideal of objective neutrality (i.e. the absence of prejudgments). The ideal of scientific neutrality reflects the defining prejudice of the Enlightenment, namely, "the

prejudice against prejudice itself'' (p. 270). Gadamer explained, ''In adopting this principle, modern science is following the rule of Cartesian doubt, accepting nothing as certain that can in any way be doubted, and adopting the idea of method that follows from this rule'' (p. 271). Gadamer saw such a portrayal of the nature of science as failing to come to grips with our own finitude and the particularity of our historical situation (p. 276). Our perceptions and understandings are shaped by personal experiences within local and temporal cultural contexts. Thus, our reasoning is not absolute, but limited and qualified.

In contrast with the Enlightenment project of elevating individual reason, Gadamer set out to rehabilitate authority and tradition. According to Gadamer, reliance upon authority is a reasonable response to our own limitations: ''It rests on acknowledgment and hence on an act of reason itself, which, aware of its own limitations, trusts to the better insight of others'' (p. 279). Authority is the source of prejudices, but these need not be false: ''...acknowledging authority is always connected with the idea that what the authority says is not irrational and arbitrary but can, in principle, be discovered to be true'' (p. 280).

Social scientists are not exempt from reliance upon tradition:

> Research in the human sciences cannot regard itself as in absolute antithesis to the way in which we, as historical beings, relate to the past. At any rate, our usual relationship to the past is not characterized by distancing and freeing ourselves from tradition. Rather, we are always situated within traditions, and this is no objectifying process – i.e. we do not conceive of what tradition says as something other, something alien. It is always part of us, a model or exemplar, a kind of cognizance that our later historical judgment would hardly regard as a kind of knowledge but as the most ingenuous affinity with tradition (p. 282).

Gadamer viewed ''unprejudiced scholarship'' as a naïve ideal that sets the social scientist apart from the rest of humanity and an inaccurate characterization of how social scientists actually do their work.

It is important to recognize that Gadamer saw traditions as dynamic, not static. A tradition must be enacted and renewed:

> The fact is that in tradition there is always an element of freedom and of history itself. Even the most genuine and pure tradition does not persist because of the inertia of what once existed. It needs to be affirmed, embraced, cultivated (p. 281).

Later, Gadamer added, ''Tradition is not simply a permanent precondition; rather, we produce it ourselves inasmuch as we understand, participate in the evolution of tradition, and hence further determine it ourselves'' (p. 293). The causality between tradition and individual behavior runs both

ways. The structures imposed by tradition constrain, but they do not eliminate individual agency (cf., Archer, 1996; Giddens, 1979). From Gadamer's perspective, fostering culture is a reasoned choice.

There are no clear rules for determining which of the prejudgments offered by tradition are correct and which are incorrect (p. 295). There is no prejudice-free perspective from which to set down procedures for evaluating prejudices. This observation is the essence of Gadamer's argument that the process of understanding cannot be reduced to a method. We always operate with a limited horizon: "The horizon is the range of vision that includes everything that can be seen from a particular vantage point" (p. 302). Progress in understanding requires that we be willing to place our prejudices at risk of refutation (p. 299). In the absence of a definitive method, the process of understanding can never be declared finished (p. 298). We can achieve new horizons, but not closure (p. 304).

Similarly, there is no method for questioning:

> The priority of the question in knowledge shows how fundamentally the idea of method is limited for knowledge, which has been the starting point for our argument as a whole. There is no such thing as a method of learning to ask questions, of learning to see what is questionable (p. 365).

Questions emerge out of dialectic interaction with the subject, as in a conversation. Questioning is an art that presupposes awareness of one's ignorance (p. 362), openness to learning (p. 366), and not being overwhelmed by the dominant opinion (p. 367).

For Gadamer, understanding was incomplete if it stopped short of application (pp. 307–324). To apply is to take action within a particular context. Applications are only partially determined by relevant prior cases. Applications are original performances that draw from past experience *and* respond to the unique particulars of the current context. There remains an on-going hermeneutical task in all applications that can never be resolved once and for all. Application is not reducible to a general method. Application cannot be fully explicit (p. 334). In contrast with an objectivist view of science, which favors detachment, application implicates oneself (pp. 334–335).

In Part III of *Truth and Method*, Gadamer described the nature and role of language. In this section, Gadamer emphasized language as the medium of understanding and for passing on traditions. Progress in understanding arises out of dialog among participants bringing distinct perspectives. The outcome of conversations can be novel understandings, not circumscribed by the prior understandings of the participants. Neither one's own

prejudices nor those of any other conversation partner are absolute. Rather, our prejudices are open to revision in the context of a dialog centered on understanding a common subject matter. The reality of the object supports the prospect of progress in understanding coming out of conversations.

Section III of Gadamer was directed more toward hermeneutics than toward the general problem of method in the social sciences that he developed earlier in his book. However, his final paragraph serves as a good summary of his thesis for the book as a whole:

> Thus there is undoubtedly no understanding that is free of all prejudices, however much the will of our knowledge must be directed toward escaping their thrall. Throughout our investigation it has emerged that the certainty achieved by using scientific methods does not suffice to guarantee truth. This especially applies to the human sciences, but it does not mean that they are less scientific; on the contrary, it justifies the claim to special human significance that they have always made. The fact that in such knowledge the knower's own being comes into play certainly shows the limits of method, but not of science. Rather what the tool of method does not achieve must – and really can – be achieved by a discipline of questioning and inquiry, a discipline that guarantees truth'' (pp. 490–491).

Gadamer's *Truth and Method* elicited critical responses from various scholars (see Grondin, 2003, pp. 297–300; Warnke, 1987, Chap. 4). Most noteworthy was the debate with Jürgen Habermas. Habermas (1977) found much to appreciate in *Truth and Method* and expressed agreement with many of its themes.[9] However, he took exception with some key points. Habermas challenged Gadamer's juxtaposing of science and hermeneutics. The natural sciences share more in common with the social sciences than Gadamer's objectivistic depiction of scientific method conveys. Both are susceptible to distortions due to researchers' personal limitations. Both require exercising human judgment. Scientific method includes interpretation and, as such, does not avoid hermeneutics. Habermas argued, ''The confrontation of 'truth' and 'method' should not have misled Gadamer to oppose hermeneutic experience abstractly to methodic knowledge as a whole...And even if it were feasible to remove the humanities entirely from the sphere of science, the sciences of action could not avoid linking empirical-analytic with hermeneutic procedures'' (1977, p. 356). Habermas' point is an important correction to Gadamer's tendency to portray the scientific method as avoiding entanglements with hermeneutics.

Habermas' (1977) broader concern was Gadamer's conservatism, as in his rehabilitation of authority and tradition. Habermas wished to shift the emphasis toward the role of critical reflection upon authority and tradition. In his words,

> Gadamer's prejudice for the rights of prejudices certified by tradition denies the power of reflection. The latter proves itself, however, in being able to reject the claim of tradition. Reflection dissolves substantially because it not only confirms, but also breaks up, dogmatic forces (Habermas, 1977, p. 358).

In the afterword to *Truth and Method*, Gadamer did not back down from his earlier position,

> I agree with Habermas that a hermeneutic fore-understanding is always in play and that it therefore requires reflexive enlightenment. But that is as far as I go with "critical rationality" because I consider perfect enlightenment illusory (2002, p. 555).

Gadamer viewed Habermas as overestimating the extent to which those engaged in dialog could throw off the constraints of prior tradition (p. 567). Rather than setting aside our perspectives, we bring them with us when we engage in reflection and dialog to arrive at new perspectives.

POLANYI ON THE PRACTICE OF SCIENCE

Michael Polanyi (1891–1976) turned to philosophy after establishing himself as a distinguished scientist working in the field of physical chemistry (Scott, 1985, pp. 2–3). He turned 55 in the year in which he published his first book of philosophical reflections on the nature of science, *Science, Faith and Society* (Polanyi, 1946). Michael Polanyi's major philosophic work was *Personal Knowledge: Towards a Post-Critical Philosophy*, first published in 1958.[10] In *Personal Knowledge*, Polanyi offered an original philosophy of science grounded in his own experience as a practitioner.

Polanyi's project was a refutation of scientific detachment. His opening chapter of *Personal Knowledge* directly challenged the presumption of objectivity. Polanyi observed that subjectivity is apparent in our choices of research topics and our criteria for accepting theories. On the choice of topics, Polanyi wrote, "...if we decided to examine the universe objectively in the sense of paying equal attention to portions of equal mass, this would result in a lifelong preoccupation with interstellar dust, relieved only at brief intervals by a survey of incandescent masses of hydrogen – not in a thousand million lifetimes would the turn come to give man even a second's notice" (p. 3). Drawing examples from mathematics and physics, Polanyi went on to argue that our criteria for accepting theories are not at all impersonal:

> We cannot truly account for our acceptance of such theories without endorsing our acknowledgment of a beauty that exhilarates and a profundity that entrances us. Yet the

> prevailing conception of science, based on the disjunction of subjectivity and objectivity, seeks – and must seek at all costs – to eliminate from science such passionate, personal, human appraisals of theories, or at least to minimize their function to that of a negligible by-play. For modern man has set up as the ideal of knowledge the conception of natural science as a set of statements which is 'objective' in the sense that its substance is entirely determined by observation, even while its presentation may be shaped by convention (pp. 15–16).

We can already see in this statement some similarities between Polanyi (1962) and Gadamer (2002; originally published in 1960), although neither cited the other.

Consider how personal judgments enter into the various steps involved in the scientific process. As just noted, the subject that we choose to study reflects personal experience and values. Prior to examining the evidence, we formulate "heuristic expectations" (p. 30). Our selection of hypotheses is inherently personal because we cannot know *a priori* the exact likelihood of confirming particular hypotheses – at least not as long as what we are attempting is a genuine discovery. Next we must make choices about what data to gather and how to evaluate their validity and reliability (p. 19). Taxonomic classification of observations relies upon tacit knowledge (pp. 348–354). When we interpret the data, our rules for inducing patterns cannot be formalized fully (p. 29).[11] Our decisions about whether the data fit a particular theory depend on traditional conventions and subjective judgments. Inducing patterns from empirical data requires *a priori* notions of what constitute patterns, as distinct from what is simply random (Chap. 3). To sum up, every step in the scientific process involves personal judgments.

Skillful performance – whether in science or other activities – involves following rules that remain unknown to the practitioner (p. 49). Words never fully capture our actions. Practitioners need not specify their rules of action to perform successfully. To focus on the rules themselves actually detracts from performance (p. 56). We can reflect upon our performances, but such analyses are distinct from the performances themselves (p. 162). When we perform, the rules that guide our performances become *subsidiary*, rather than *focal* (pp. 55–57). Thus, despite formalization of aspects of our scientific practices, we never eliminate the personal role of the investigator.

The distinction between focal and subsidiary awareness underlies Polanyi's notion of "indwelling." He used this term in a variety of contexts, including our relationships to our own bodies (p. 58), use of tools (p. 59), assimilating and transmitting a culture (pp. 173–174), and thinking within the conceptual frameworks of mathematics, art, and religion (p. 283). In

each case, we so identify with that in which we dwell that it forms a part of ourselves and how we experience phenomena outside ourselves. In other words, that in which we dwell becomes subsidiary as we attend focally to some object of study or experience. Polanyi's example of how we use tools clarifies indwelling:

> Our subsidiary awareness of tools and probes can be regarded now as the act of making them form a part of our own body. The way we use a hammer or a blind man uses his stick, shows in fact that in both cases we shift outwards the points at which we make contact with the things that we observe as objects outside ourselves. While we rely on a tool or a probe, these are not handled as external objects. We may test the tool for its effectiveness or the probe for its suitability, e.g. in discovering the hidden details of a cavity, but the tool and the probe can never lie in the field of these operations; they remain necessarily on our side of it, forming part of ourselves, the operating persons. We pour ourselves out into them and assimilate them as parts of our own existence. We accept them existentially by dwelling in them (p. 59).

That in which we dwell shapes how we experience the world and interpret the data that we take in. It follows that our relationships to theories and research methods can also be described as acts of indwelling.

Because the requisite knowledge for practice always has a tacit dimension, we cannot depersonalize the transfer of skills. Skills are passed on from masters to apprentices and, as such, their acquisition involves submission to authority and tradition. Polanyi recognized,

> To learn by example is to submit to authority. You follow your master because you trust his manner of doing things even when you cannot analyse and account in detail for its effectiveness. By watching the master and emulating his efforts in the presence of his example, the apprentice unconsciously picks up the rules of the art, including those which are not explicitly known to the master himself. These hidden rules can be assimilated only by a person who surrenders himself to that extent uncritically to the imitation of another. A society which wants to preserve a fund of personal knowledge must submit to tradition (p. 53).

Similarly, the acquisition of tastes, as expressed in connoisseurship, requires learning from others, not just following explicit rules (p. 54).

The interdependencies within a community of shared knowledge go beyond the mentor–apprentice relationship. Because the scope of individual experience is so limited, we must rely upon the findings of others: "...a system of scientific facts and standards can be said to exist only to the extent to which each scientist trusts all the others, to uphold his own special sector of the system in respect of his research, his teaching and his administrative actions" (p. 375). Such trust is another instance of our submission to authority; in this case, the authority is mutual among peers. Our interactions within a scientific community recognize traditions as authoritative, but not

final (p. 164). All theories are amenable to change. Despite granting our respect, we admit the fallibility of even the greatest scholars. As Polanyi observed, "Every acceptance of authority is qualified by some measure of reaction to it or even against it" (p. 208).

Progress in knowledge involves (1) interpretation based upon pre-existing schema and (2) revisions of those schema as we encounter novel situations. The process of adaptive learning brings together pre-existing schema, personal judgment, and the things themselves (pp. 103–104). This is quite similar to Gadamer's description of the hermeneutical task. Like Gadamer, Polanyi held to realism as a basis for the possibility of progress in knowledge over time: "The modification of our intellectual identity is entered upon in the hope of achieving thereby closer contact with reality" (p. 106).

Polanyi shared Gadamer's conviction that the steps to progress in knowledge are not reducible to an impersonal method (pp. 370–371). As he stated, "...no solution of a problem can be accredited as a discovery if it is achieved by a procedure following definite rules" (p. 123). One must exercise personal judgment in the context of novel situations. Novel situations present perplexing problems that we seek to resolve through further exploration of the data for clues to the unknown (p. 120, pp. 127–128). The search for a solution involves two operations: "We must (1) set out the problem in suitable symbols and continuously reorganize its representation with a view to eliciting some new suggestive aspects of it, and concurrently (2) ransack our memory for any similar problem of which the solution is known" (p. 128).

Because the criteria for knowledge are esthetic and personal, science is a passionate activity (p. 133). Polanyi saw such passion as essential to our intellectual heritage (p. 134). Passion sustains scientific activities through years of labor (p. 143). Our emotions are not just byproducts of scientific discoveries; rather, they guide and energize us during the process of discovery. Polanyi argued:

> Theories of the scientific method which try to explain the establishment of scientific truth by any purely objective formal procedure are doomed to failure. Any process of enquiry unguided by intellectual passions would inevitably spread out into a desert of trivialities. Our vision of reality, to which our sense of scientific beauty responds, must suggest to us the kind of question that it should be reasonable and interesting to explore. It should recommend the kind of conceptions and empirical relations that are intrinsically plausible and which should therefore be upheld, even when some evidence seems to contradict them, and tell us also, on the other hand, what empirical connections to reject as specious, even though there is evidence for them – evidence that we may as yet be unable to account for on any other assumptions. In fact, without a scale of interest and plau-

> sibility based on a vision of reality, nothing can be discovered that is of value to science; and only our grasp of scientific beauty, responding to the evidence of our senses, can evoke this vision (p. 135).

Scientific discoveries should be *interesting*, as well as accurate and significant (pp. 135–136).[12]

Polanyi indicated a threefold role of passions in science (p. 159). First, passions serve a *selective* function guiding the choice of subject. Second, they have a *heuristic* function that "...links our appreciation of scientific value to a vision of reality, which serves as a guide to enquiry" (p. 159). Third, they have a *persuasive* function that sustains scientific debates. Such debates occur despite the obstacles posed by seemingly incommensurable paradigms (p. 151).

Polanyi rejected pragmatism as a philosophy of science. As already noted, Polanyi saw achieving contact with reality as the goal of science – "...a contact destined to reveal itself further by an indefinite range of yet unforeseen consequences" (p. 147). We never know fully the implications of our discoveries: "The mark of true discovery is not its fruitfulness but the *intimation* of its fruitfulness" (p. 148). Furthermore, evaluating the fruitfulness of a discovery requires some additional criterion (or criteria) that cannot ultimately be justified pragmatically. Polanyi argued that strict pragmatism undermines science by driving out passionate inquiry by specialists seeking truth for its own sake.

Science takes place within a community. Scientists do not work alone; they rely upon others to cultivate and sustain their intellectual passions (p. 203). Polanyi stressed the importance of "conviviality" among scientists (Chap. 7). The values that uphold science must be supported not only by fellow scientists, but by society as a whole (see also Polanyi, 1946). Polanyi summarized, "The recognition granted in a free society to the independent growth of science, art and morality, involves a dedication of society to the fostering of a specific tradition of thought, transmitted and cultivated by a particular group of authoritative specialists, perpetuating themselves by co-option" (p. 244).

Polanyi's epistemology can be described as a refutation of doubt as the starting point of knowledge, as proposed by Descartes, and affirmation of the necessity of faith. Polanyi explained:

> The method of doubt is a logical corollary of objectivism. It trusts that the uprooting of all voluntary components of belief will leave behind unassailed a residue of knowledge that is completely determined by the objective evidence. Critical thought trusted this method unconditionally for avoiding error and establishing truth (p. 269).

Polanyi's use of the verb "trust" in this explanation points toward the direction of his critique of doubt as the starting point for knowledge. He argued that it is impossible to suspend all of our beliefs and thereby establish an indubitable foundation for knowledge: "...to claim that you strictly refrain from believing anything that could be disproved is merely to cloak your own will to believe your beliefs behind a false pretence of self-critical severity" (p. 271). We cannot suspend all beliefs simultaneously (p. 294). We are limited to doubting some of our beliefs on the basis of other beliefs that we continue to hold. Polanyi concluded, "Objectivism has totally falsified our conception of truth, by exalting what we can know and prove, while covering up with ambiguous utterances all that we know and *cannot* prove, even though the latter knowledge underlies, and must ultimately set its seal to, all that we *can* prove" (p. 286).

Thus, knowledge depends not upon achieving impersonal objectivity, but upon fiduciary commitments. Polanyi stated his own fundamental commitment in the following way: "I believe that in spite of the hazards involved, I am called upon to search for the truth and state my findings" (p. 299). We commit to that which we believe to be true and, therefore, binding upon ourselves and others (Chap. 10). For Polanyi, the ethical imperative of submission to truth arises from the desire for knowledge. Apart from personal commitments, which express faith, there can be no knowledge. We avoid complete subjectivity by our belief that we have made contact with some knowable aspect of reality (p. 311). As Polanyi's statement indicates, our knowledge commitments are risky: "The possibility of error is a necessary element of any belief bearing on reality, and to withhold belief on the grounds of such a hazard is to break off all contact with reality" (p. 315).

IMPLICATIONS FOR MANAGEMENT RESEARCH

Critical Realism

Gadamer was interested primarily in the problem of method in hermeneutics and the social sciences. Polanyi sought to present an alternative to the norm of impersonal objectivity in science. Despite their distinct research objectives, their perspectives are quite compatible. Together, they offer important insights into the personal and interpersonal dimensions of scientific methods and practices. Their thinking also converged on a common philosophy of science. Although their work predated the label "critical realism," the

underlying assumptions of Gadamer's hermeneutics and Polanyi's explanation of scientific practice fit well within this philosophy of science.

Roy Bhaskar initiated critical realism with the publication of *A Realist theory of Science* in 1975.[13] In his opening chapter, Bhaskar argued that an explanation of knowledge must recognize "...both (1) the social character of science and (2) the independence from science of the objects of scientific thought" (1978, p. 24). These two points summarize the essentials of the critical realist philosophy of science. They also convey key assumptions held by Gadamer and Polanyi. The first claim reflects an appreciation that scientific knowledge is a social product. The second posits an ontology that makes science possible. Critical realists hold that there is more to reality than we currently know and more than what we as researchers create. As such, discoveries that advance knowledge are possible.

Postulating unobservables is not problematic for critical realists (Harré, 1970), as it was for logical positivists. Logical positivists, such as Ayer (1952) and Carnap (1974), rejected claims about a transcendent reality. Instead, they asserted empirical verification as the essential criterion of science. From the logical positivist perspective, it is senseless to include unobservables within science. By contrast, critical realists allow that constructs and the mechanisms that relate constructs to one another may be unobservable, yet nonetheless relevant to scientific theorizing. This allowance for unobservables comes from a recognition of the limitations of our human senses and experiences. Admitting unobservables is in keeping with nearly all of the theorizing within strategic management. Godfrey and Hill (1995) argued for realism, and against positivism, in strategic management research. They pointed out that key constructs within transaction cost theory, agency theory, and the resource-based view of the firm are unobservable.

Like Gadamer and Polanyi, Bhaskar (1978) presented ontological realism as fundamental to science. Bhaskar explained,

> For it is only if the working scientist possesses the concept of an ontological realm, distinct from his current claims to knowledge of it, that he can philosophically think out the possibility of a rational criticism of these claims. To be a fallibilist about knowledge, it is necessary to be a realist about things (1978, p. 43).

Knowledge depends upon the existence of the things themselves, not vice versa.

Nevertheless, realism is not incompatible with social construction. Berger and Luckmann's (1967) theory of social construction was not antirealist. They denied neither the material aspects of social phenomena nor the reality of social phenomena. Even Weick (1979) is not antirealist; he approaches

organizations and their enacted environments as quite real (Sandelands & Drazin, 1989). If realism meant denying socially constructed phenomena, then it would provide no basis for studying organizations. The compatibility of realism and social construction can be seen in the advances of critical realist research within sociology (e.g. Archer, 1995).

The exchange between Mir and Watson (2000, 2001) and Kwan and Tsang (2001) in *Strategic Management Journal* pointed out the extent to which the constructivist and critical realist perspectives overlap. Mir and Watson's (2000) characterization of constructivism was, for the most part, compatible with critical realism. They summarized constructivism as holding to a view of research in all of its aspects – questions, procedures, and evidence – as theory-dependent, rather than objective. Their key point of contention with Kwan and Tsang proved to be the extent to which researchers and the phenomena they study are interdependent. At issue was whether social scientists create theories of independent phenomena or create the social phenomena themselves. Posing the issue in this way overlooks the possibility that the influence of researchers may often lie somewhere between these two extremes. As researchers, we should consider the extent to which the phenomena that we study are dependent upon our actions and the discourse among ourselves. As researchers, our actions and discourse may occur largely outside the social phenomena that we study, or they may not. On one hand, our laments about our lack of influence on practice (e.g. Hambrick, 1994) may often be warranted, which frees us up to study organizational phenomena as largely independent of ourselves. On the other hand, some of our research involves direct interventions that transform organizations. Not all research needs to be action research (see Greenwood & Levin, 1998), but some of it can (and should) be.

Bradbury and Lichtenstein (2000) called management researchers to consider our relationships to the people who inhabit the organizations that we study. They proposed a "relationality orientation" to organizational research:

> ...a relational theorist will be conscious of the impact of her/his research on what is being researched, and too on how that research impacts her/himself. Such a scholar enters an organization as if it were an extended set of relationships. S/he thereby places more attention on the "space between" – the space between subject and object, subject and research, researcher and subject, and the reflexivity of the research process itself (Bradbury & Lichtenstein, 2000, p. 551).

Bradbury and Lichtenstein distinguished between *multiperson* research that takes place within a community of scholars, *interpersonal* research that is done together with the research subject, and *personal* research done by and

on oneself. Multiperson research can be done in ways that have little effect on the phenomena of interest. Interpersonal research requires greater involvement of the subjects in designing the study, providing data, and interpreting results. The goals of interpersonal research must align with the interests of the subject. Personal research emphasizes usefulness over generalizability.

As a philosophical position, realism is best understood in contrast with solipsism. Solipsism entails either a denial that reality outside oneself can be known (weak solipsism), or an ontological claim that there is no reality outside oneself (strong solipsism). It is not difficult to figure out where most management researchers stand on ontological realism versus solipsism. Although we are rarely explicit about our commitment to realism, it is frequently apparent in management and organization research (Ackroyd & Fleetwood, 2000; Fleetwood & Ackroyd, 2004).

In addition to questions about whether social construction fits within realism, another concern is whether realism applies to social phenomena that are dynamic, not static. Like the question of social construction, this is not a serious challenge to realism. We simply conceptualize the real in terms of its dynamics. Chia (1997) contrasted an "ontology of being" and a "becoming ontology." An ontology of being is one "...in which the 'thingness' of things, social entities, and their properties and attributes are taken to be more fundamentally real than actions, interactions and relationships" (Chia, 1997, p. 690). A becoming ontology privileges "activity and movement" over "substance and entities" (Chia, 1997, p. 696). I would argue that there is no compelling case for excluding "actions" or "things" from ontology. Ontology must encompass both. The claim that "change is pervasive and indivisible" (Tsoukas & Chia, 2002, p. 569) is a claim about *something*; thus, it is consistent with realism. We need not posit a static world, nor do we need to privilege "activity and movement" above "substance and entities." Ontology can encompass things in action and transition.

Methodological Pluralism

Gadamer and Polanyi contended that presumptions are unavoidable (see also Rescher, 1988). Gadamer insisted on approaching the problem of method from the recognition of our own finitude and the particularity of our place in history. The creation of distinct traditions and our reliance upon them are necessary implications of our human limitations. Polanyi held that knowledge arises within communities of shared commitments and

practices. Communities accumulate distinct knowledge because of (1) the fiduciary aspect of all knowledge and (2) the time and effort required to disseminate knowledge. Dissemination of tacit knowledge occurs only through interpersonal relationships.

These observations help explain why we see such a diverse array of theories and methods within management research. Our field attracts researchers with varied disciplinary backgrounds. We have all developed commitments based upon our past educational and professional experiences. As Astley (1985, p. 505) observed, "...different theorists bring different intentions to the study of administration; they are set on investigating quite different things, interpreting reality through their own conceptual filters, and drawing conclusions that fit their own world views." From a realist perspective, we would point to the multifaceted nature of organizations and the distinct phenomena that we study as further reasons for the pluralism within our field (see McKelvey, 1997). Both traditions and the multifaceted nature of our subject account for the plurality of methods in management research.

Methodological pluralism poses some very practical challenges. As researchers and reviewers, we face the dual demands of deepening our knowledge of particular methods and gaining knowledge of diverse methods. As writers seeking to publish, we face formidable challenges because the methods that we choose may be unfamiliar to reviewers and many of our readers. We tend to confer scholarly recognition on those who become proficient in particular methods and advance the state of practice within their specialization. Although knowledge of diverse methods is important to reading and evaluating the research within our field, gaining such knowledge has a high opportunity cost. At its extreme, pluralism poses the problem of incommensurable methods and knowledge claims (Scherer, 1998).

Discussions of pluralism within our field have tended to emphasize scientific paradigms (see Kuhn, 1970) rather than methods. Pfeffer's controversial (1993) case for paradigm consensus triggered a high-profile debate regarding pluralism in organization studies. Drawing from prior research across academic fields, Pfeffer noted that paradigm consensus is associated with more rapid advancement in knowledge and greater ability to attract resources for research than paradigm dissension. He reasoned that consensus was, at least in part, a choice and that such a choice was needed to correct for excessive theoretical and methodological pluralism.

Pfeffer's prescription incited critical responses from various authors. Perrow's (1994, p. 192) critique was based on ontology: "There can be no paradigm for all organizations or all time because organizations are

ever-evolving responses to social change, and thus the context of organizational behavior is a major variable.'' Cannella and Paetzold's rebuttal was based on epistemology: ''Because we find ourselves unable to determine how close our theories are to some absolute truth, we are unable to evaluate paradigms in a way that would enable us to know that any particular one is *a priori* deserving of a dominant position in organization science'' (1994, p. 332). Van Maanen's (1995b) rejoinder was based on hermeneutics; he called attention to the indeterminate nature of research texts and the personal nature of textual interpretation. Textual indeterminacy precludes definitive judgments about the merits of alternative research paradigms.[14]

Management researchers addressing research methods tend to advocate either (1) innovative methods to overcome the shortcomings of conventional methods or (2) diversity in methods to overcome the shortcomings of any given approach – traditional or innovative. Daft and Lewin (1990) encouraged ''heretical research methods'' as a way to ''break out of the normal science straightjacket.'' They explained,

> The point of heretical research methods is to find new channels through which to obtain organizational insights and to change the mix of research methods. Although no method is truly heretical, researchers should be encouraged to do whatever it takes to learn about organizations (Daft & Lewin, 1990, p. 6).

Two examples that they provided were examining outliers and doing longitudinal field research. Also reacting against the ''normal science straightjacket,'' Bettis (1991, p. 318) advocated, ''In addition to large sample, cross-sectional, multivariate statistical studies, encourage qualitative studies, longitudinal studies, studies that cross levels of analysis, studies of outliers, mathematical modeling, speculative studies, simulations, laboratory experiments, case studies, interviews, and histories of firms, industries, businesses, innovations, and managers.''

Lee (1991) promoted combining both positivist and interpretive approaches to organizational research. Positivist approaches seek to emulate the natural sciences through approaches such as ''...inferential statistics, hypothesis testing, mathematical analysis, and experimental and quasi-experimental design'' (Lee, 1991, p. 342). Interpretive approaches include procedures associated with ''...ethnography, hermeneutics, phenomenology, and case studies'' (Lee, 1991, p. 342). Lee viewed positivist and interpretive approaches as informing one another. Thus, he advocated iterating between the two approaches.

Gioia and Pitre (1990) advocated a multiparadigm approach to theory building. They argued that because different approaches involve distinct

ontological and epistemological assumptions, they provide complementary perspectives from which to understand organizations. They recognized that the needed – and challenging – work to be done involves triangulating across paradigms. Building on this perspective, Lewis and Grimes (1999, p. 676) elaborated a process of multiparadigm theory building through "metatriangulation," which they defined as "a process of building theory from multiple paradigms roughly analogous to traditional (i.e. single-paradigm) triangulation."

Claims about whether we have too many or too few paradigms or methods are premature. Such determinations could only be made once we have completed the research and have understood the phenomena of interest, but such closure in knowledge is an unattainable ideal. While we are still in the midst of the uncertainty of the process of discovery, we can make no such judgments. What we can say is that pluralism produces richer understandings, but it also slows down the learning process and defers knowledge exploitation (March, 1991).

Pluralism within management research leads us to consider hermeneutics. From Gadamer, we understand that methods reflect traditions. Our ways of doing science can never be free of prejudices (see also Ricoeur, 1981, Chap. 9). There is no objective method. Furthermore, there is no prejudice-free meta-method for mediating between particular methods. We never can step outside our personal limitations when we engage in research or read the work of others. Both the writer and the reader operate from limited perspectives. In Gadamer's terminology, we each have our own "horizon." Hermeneutics grapples with the challenges of learning across distinct horizons. For Gadamer, this meant seeking a "fusion of horizons" providing a novel perspective not fully circumscribed by its predecessors.

Research on the sociology of scientific knowledge reveals the persistence of prejudices in research traditions. Kuhn (1970) documented the continuity of "normal science" – i.e. incremental and conservative puzzle-solving – even in the face of accumulated anomalies. Lakatos (1970) portrayed scientific research programs as operating with both "negative heuristics" – telling researchers which paths to avoid – and "positive heuristics" affirming which paths to pursue. Laudan described "research traditions" as exhibiting metaphysical and methodological commitments: "Put simplistically, a *research tradition is thus a set of ontological and methodological 'do's' and 'don'ts*" (1977, p. 80; emphasis in original). Although there were differences in the theories put forward by Lakatos and Laudan (see Laudan, 1977, pp. 76–78), they both affirmed that untestable core beliefs and methods are

more basic than theories for explaining the persistence of research traditions.

Although our prejudices confine our thinking and methods, they need not be absolute. Gadamer saw prospects for resolving the hermeneutical impasse in the recognition that, despite their rigidities, horizons are not fixed:

> Just as the individual is never simply an individual because he is always in understanding with others, so too the closed horizon that is supposed to enclose a culture is an abstraction. The historical movement of human life consists in the fact that it is never absolutely bound to any one standpoint, and hence can never have a truly closed horizon (Gadamer, 2002, p. 304).

Hence, by acknowledging the distinct horizons that we bring to our research, we also see the prospect for advancing knowledge through reading outside our research traditions and engaging in dialog across traditions. Hermeneutics is not a universal method, but a practice of dialectic learning through reading and conversation (Schrag, 1992).

This discussion of Gadamer's *Truth and Method* points to the potential value for management researchers of the literature on hermeneutics. In addition to Gadamer's writings, management researchers interested in exploring hermeneutics should consider Ricoeur's (1981) *Hermeneutics and the Human Sciences*, Bernstein's (1983) *Beyond objectivism and relativism: Science, hermeneutics, and praxis*, and Habermas' (1984) *The Theory of Communicative Action*. Hermeneutics informs the problems and possibilities of pluralism, and the processes involved in advancing knowledge. Hermeneutics prioritizes dialog over any particular method. Hermeneutics advocates practices consistent with Mahoney's (1993, p. 188) advice that, "...strategy research should concern itself with continuing the conversation of the field rather than insisting upon a place for universal methodological criteria within that conversation."

The use of specialized methods can provide unique insights into the phenomena of interest to management researchers, but they also can make research inaccessible to those who lack the requisite expertise. Methods impose entry and mobility barriers. They can shut other academics and practitioners out of our conversations. The more knowledge and skill needed to use a method, the fewer are those who can participate as contributors and readers. The point of these observations is not that we should avoid specialized methods. Diverse methods may be essential to advance management knowledge. While employing a diverse array of methods, we should seek to make our research accessible to outsiders. Our shared hermeneutical

task is to find ways in which pluralism can advance knowledge without balkanizing the field. Bradbury and Lichtenstein stated the goal in this way:

> In a sense, the very tone used to write up one's research can increase or decrease interpersonal and professional dialogue on a topic. From a relationality perspective, the criteria of promoting intelligibility through one's research puts an ethical value on presenting one's findings in a way that allows others to generate alternative interpretations. This move serves to increase communication, creativity, and validity in science in the long term (2000, p. 561).

In a field with limited resources for this task, the literature on hermeneutics offers background to motivate and guide our communication.

We also must consider the problems of sustaining dialog among researchers and practitioners given the economic interests at stake in management research. Can a field centered on "competitive advantage" sustain open conversation? In other words, does the normative status afforded competitive advantage drive out the exchange of ideas needed to advance knowledge? The writings of Gadamer and Polanyi challenge us to consider whether the norms and incentives within our field are consistent with advancing knowledge. Our own theories tell us that the production of knowledge always involves a tension between self-interest and open conversation. Fabian (2000) suggested that our field is less about conversation than it is about "impassioned debate." Our participation is never disinterested: "Although the debate is complex and, yes, can be tiresome, we need to take the debate personally, because in the end it is as personal as it can be: it is our science and our profession" (Fabian, 2000, p. 367). Polanyi encouraged scientists to be both passionate in their individual pursuits and committed to advancing the collective knowledge of a research community. The paradox of seeking knowledge-based competitive advantages is that this goal may drive out the dialog necessary to advance the collective knowledge of our field.

Practice and Relationality

The writings of Gadamer and Polanyi reflect two important shifts within philosophical discussions that became prominent during the latter half of the twentieth century. The first of these was the *practice turn*. This was an epistemological shift away from knowledge as theory to knowledge as exhibited in performance (see Bernstein, 1983). Gadamer understood hermeneutics as a practice not reducible to a method. Drawing from his own experience, Polanyi grounded his philosophy of science in a description of what scientists actually do, i.e. their practices. The second philosophical

shift was the *turn toward relationality* (Shults, 2003). With this shift, the individual is decentered (Schrag, 1986). Community replaces the individual as the locus of rationality, knowledge, and ethics. Gadamer highlighted participation in community as essential to hermeneutics and social science. Polanyi emphasized the communal aspects of conviviality and mutual accountability in science.

These two shifts are becoming increasingly evident in the content of management research. Research on organizational action (Heracleous, 2003; Jarzabkowski, 2003; Johnson, Melin, & Whittington, 2003; Tsoukas & Knudsen, 2002) and strategy as practice (Hendry, 2000; Whittington, 1996, 2002) reflect the practice turn. Research on communities of practice emphasizes both practices and relationality. The relational emphasis in this research is apparent in its focus on the community as a relevant level of analysis for understanding learning processes (Lee & Cole, 2003). Practices form, evolve, and disseminate within communities (Brown & Duguid, 1991, 2001; Cohendet & Llerena, 2003).

Although the themes of practice and relationality are gaining greater attention in the content of our theories, we have not yet given much attention to their implications for how we understand our own research activities and methods. For example, Brown and Duguid's (1991) threefold summary of work practices – narration, collaboration, and social construction – appear to be quite consistent with what we as researchers do. First, *narration* occurs in response to problems. When confronted with a problem, we tell stories to organize information and generate and test hypotheses. Narratives also serve as a memory device to preserve practical experiences for future recall. Second, *collaboration* is essential to learning: "Not only is the learning in this case inseparable from working, but also individual learning is inseparable from collective learning. The insight accumulated is not a private substance, but socially constructed and distributed" (Brown & Duguid, 1991, p. 46). Third, *social construction* is integrally related to narration and collaboration, and is evident in the creation of shared understandings and identities. Thus, research within our own field on learning within communities challenges our emphasis on individual – as opposed to communal – contributions to theory building.

Practices, even those of science, are always more than can be captured as codified methods. Practices have an inherently personal dimension. Polanyi explained:

> Admittedly, there are rules which give valuable guidance to scientific discovery, but they are merely *rules of art*. The application of rules must always rely ultimately on acts not determined by rule. Such acts may be fairly obvious, in which case the rule is said to be

> precise. But to produce an object by following a precise prescription is a process of manufacture and not the creation of a work of art. And likewise, to acquire new knowledge by a prescribed manipulation is to make a survey and not a discovery. The rules of scientific enquiry leave their own application wide open, to be decided by the scientist's judgement. This is his major function. It includes the finding of a good problem, and of the surmises to pursue it, and the recognition of a discovery that solves it. In each decision the scientist may rely on the support of a rule; but he is then selecting a rule that applies to the case, much as the golfer chooses a suitable club for his next stroke.
>
> Viewed from outside, as I have just described him, the scientist may appear as a mere truth-finding machine steered by intuitive sensibility. But this view overlooks the curious fact that from beginning to end he is himself the ultimate judge in deciding on each consecutive step of his enquiry. He has to arbitrate all the time between his own passionate intuition and his own critical restraint of it (1946, pp. 14–15).

Among organizational researchers, Pettigrew (1990, pp. 285–286) made some related observations:

> ...research is a craft activity. It is not just the application of a formal set of techniques and rules. A craft activity involves the application of skills, knowledge, and the person in varying settings. Within these settings individual judgments are made in the context of a wider system of collective rules and communication. Even if the methods advocated in this paper had all been codified and written down (and there is much productive work to do in this direction) craft skills would still be required to interpret and apply such codifications according to the particular nuances and subtleties of each research project and site. Strauss (1987) is very clear on this, there are no simple general rules which can be applied in a standardised fashion.

In keeping with this perspective, Weick (1989, p. 707) summarized, "Whatever strategy is used, there will always be an uncodifiable step that relies on the insight and imagination of the researcher." Acknowledging personal judgment and tacit knowledge does not undermine science; rather, such a description provides a basis for determining feasible norms of scientific practice.

We develop personal judgment and gain access to the tacit dimension of knowledge through working together with others more skilled than ourselves. Polanyi (1962, Chap. 4) pointed to mentoring relationships as essential for the transmission of skills, traditions, and connoisseurship. According to Polanyi, learning is inherently a communal activity. The literature on communities of practice affirms this contention. Lave and Wenger (1991) and Brown and Duguid (1991) emphasized "legitimate peripheral participation" (LPP) as essential to learning:

> Learning, from the viewpoint of LPP, essentially involves becoming an "insider." Learners do not receive or even construct abstract, "objective," individual knowledge; rather, they learn to function in a community – be it a community of nuclear physicists,

> cabinet makers, high school classmates, street-corner society, or, as in the case under study, service technicians. They acquire that particular community's subjective viewpoint and learn to speak its language. In short, they are enculturated (Brown, Collins, and Duguid 1989). Learners are acquiring not explicit, formal "expert knowledge," but the embodied ability to behave as community members (Brown & Duguid, 1991, p. 48).

Learning in community involves transmission of narratives and tacit knowledge, and formation of identity (Polanyi, 1962; Brown & Duguid, 1991). By implication, codified presentations of research methods convey only emaciated views of what research actually entails.

Research on the sociology of scientific knowledge has followed Polanyi's view of science as social practices (e.g. Pickering, 1992; Rouse, 1996, Chap. 5). Knorr Cetina (1999) described scientific communities as "epistemic cultures," i.e. cultures that generate and legitimate knowledge claims. She portrayed scientific practices as reflecting distinct local cultures. Her work highlighted the fragmentation of science into subcultures working from distinct epistemic criteria. This fragmentation is evident not only in the bifurcation between the natural sciences and social sciences, as portrayed by Gadamer, but also in the epistemic diversity within each of these categories.

For Polanyi, the observation that science takes place within a community carried important ethical implications. Progress in science requires personal commitments to a scientific community and the broader society (Polanyi, 1946). Polanyi (1962, Chap. 7) emphasized the qualitative aspects of interpersonal relationships within a scientific community in his discussion of "conviviality." Conviviality reinforces essential mutual commitments to integrity and trust in the creation and transmission of knowledge. Because of the specialized and tacit dimensions of scientific practices, society's monitoring of scientists is necessarily incomplete. As such, scientists must hold themselves personally accountable for their own knowledge claims (Polanyi, 1969, Chap. 5). Our claims to knowledge are made with "universal intent," which calls for submission of ourselves and others to their truth.

Such commitments are inconsistent with objectivity as normative for science (Polanyi, 1969, Chap. 3). Rather, ethical commitments grow out of one's identification with a larger community (see MacIntyre, 1984). Polanyi (1962) used the term "indwelling" to express such self-defining commitments. To indwell beliefs means that the beliefs shape one's identity, understanding, and behaviors. Our claims to knowledge ultimately depend upon beliefs that we affirm uncritically. As noted earlier in the discussion

of Polanyi's fiduciary perspective, knowledge claims always entail risk and relationships with others. Risk arises because of our fallibility. The relational aspect is our commitment that our knowledge claims are binding upon ourselves as well as others (Polanyi, 1962, Chap. 10). Following Gadamer and Polanyi, our knowledge can never be objective or final, but it can, nevertheless, be true.

DISCUSSION

This chapter has explored the human side of research methods. Following Gadamer and Polanyi, I have argued that objectivistic portrayals of research methods omit essential elements of our participation as researchers. Our exercising of judgment and reliance upon tacit knowledge are unavoidable aspects of the research process. As such, research has an inherently personal dimension. Research is also inherently interpersonal. We draw upon tradition and knowledge dispersed within a community. We look to others for validation of our findings and their meaning. Acknowledging others as experts implies trusting submission to authority. This final section discusses the contributions of this chapter and some specific implications of the personalized account of research methods offered here.

I have not encountered any prior research comparing the writings of Hans-Georg Gadamer and Michael Polanyi. As such, one of the intended contributions of this study was to show how their writings coincide on many points. This is a somewhat surprising finding given the disparate questions of interest – Gadamer primarily was interested in hermeneutics and the social sciences, while Polanyi focused on the natural sciences – and the absence of acknowledgments of each other's work in their writings. Their common interest in refuting objectivism and their shared experience in the academic milieu of twentieth-century Germany may account for the complementary themes in their writings.

By providing an overview of the thinking of Gadamer and Polanyi on research methods, I hope to encourage management researchers to read the works of these two philosophers and grapple with their implications for our field. Gadamer provides a helpful starting point for considering how the literature on hermeneutics could inform the process of learning within a historically situated scholarly community. As others have charged (Brown & Duguid, 2001; Foss, 2003; Tsoukas, 2003), for too long, we have been content to cite Polanyi for the concept of "tacit knowledge" without carefully

reading his writings. Polanyi's work can inform our understanding of research practices as well as organizational knowledge.

Gadamer and Polanyi preferred that we recognize our human limitations, rather than hold our research efforts to an unattainable ideal of impersonal objectivity or, worse yet, falsely portray our work as having attained to such an ideal. Gadamer sought to rehabilitate tradition and authority as bases for understanding within the social sciences and hermeneutics. He argued forcefully against the Enlightenment's "prejudice against prejudice." Polanyi refuted scientific detachment as expressed in logical positivism. In its place, he offered a portrayal of science as personal and communal. Science is more than methods and theories, it requires skillful practices, which are inherently personal because they express tacit knowledge.

In rejecting the scientific method as absolute, Gadamer offered a balanced alternative perspective reflecting dependence upon tradition – as an acknowledgment of our personal limitations – and allowance for critical reflection – because methods are never final or absolute. By contrast, Feyerabend (1975) advocated methodological anarchy; he espoused *rejection* as a normative approach to traditional methods. There are a variety of problems with Feyerabend's logic. It does not follow from *specific* historical cases of scientific progress resulting from breaking with established methods that *all* methods are flawed (Laudan, 1996, Chap. 5). His position is self-refuting (i.e. applying Feyerabend's rejection of methods to his own perspective would require adherence to tradition methods). It also overlooks past progress in knowledge using conventional research methods, and the role of anomalies produced by normal science in bringing about paradigm shifts (Kuhn, 1970). Although breaking with conventional methods may be beneficial to our collective learning over the long-run, failure to follow prevailing conventions often has adverse immediate consequences – particularly for the rebel (March, 1991). Gadamer advocated a position between the extremes of strict adherence to conventions and anarchism. We can challenge our methods selectively and sequentially, but we can never do so simultaneously and universally.

In keeping with his rejection of scientific detachment, Polanyi acknowledged and affirmed the role of personal passion in guiding and sustaining research. Polanyi explained his experience as a scientist as passionate pursuit of truth. Our research subjects should capture our attention because they offer a context for expressing our personal and collective values. Those of us who have come to view our research as merely instrumental to our career objectives need to get back in touch with our concerns and convictions about organizations and their roles in our lives. Such passions guided us into

this field of research, but can become dull as we shift our focus to the extrinsic pressures and rewards of academic work.

If we care about the role of organizations in peoples' lives, objectivity is neither feasible nor normative for our research. Rather, we adopt a "relationality orientation" (Bradbury & Lichtenstein, 2000). As argued earlier, the phenomena that we study are rarely either completely independent or completely dependent upon ourselves as researchers. Our research results in varying degrees of interdependence with the people and organizations that we study. Research methods differ in their degree of obtrusiveness (Argyris, 1968; Webb & Weick, 1979). We should be attentive to how our actions and dialog as researchers influence the phenomena that we study. However, we should not shy away from making a difference in organizations. Research cannot be a fully impassioned activity if we prohibited ourselves from changing organizations in ways that reflect our personal and shared values. Distancing ourselves from organizational phenomena not only diminishes our relevance, it also causes us to miss opportunities to advance our understanding by observing organizations firsthand. If application is an essential part of knowledge, as Gadamer contended, then we must continue to challenge ourselves to move beyond understanding as codified theorizing, to understanding as practical performance.

Our beliefs do not derive from indubitable foundations (Bonjour, 1978). Because of its tacit dimension, knowledge can never be fully explicit. Knowledge claims reflect personal commitments to presumptive beliefs (Rescher, 1988). The epistemic criteria that we use to judge knowledge claims are fallible. They emerge within local social contexts and evolve over time. Polanyi (1962, Chap. 9) recognized that Descartes' program of universal doubt was an impossible starting point for arriving at knowledge. He argued that faith is essential to uphold the practices of science and claims to knowledge. Because of our personal and collective fallibility, knowledge claims involve risk. The natural sciences and social sciences are alike in this regard. As a fiduciary activity, science cannot avoid ethical and theological considerations (Stenmark, 1995; van Huyssteen, 1999). The fact-value dichotomy of logical positivism is no longer sustainable if science is a fiduciary activity.

For those of us trained – explicitly or implicitly – to aspire to objectivity, acknowledging the personal aspects of research methods can be disconcerting. If objectivism and relativism were the only alternatives, then it would seem that abandoning objectivism undermines the possibility of science. However, Gadamer and Polanyi did not see objectivism and relativism as the only alternatives. They argued for another alternative that recognizes

our fallibility and lack of objectivity, yet remains optimistic about progress in knowledge. Their commitment to ontological realism was essential to their optimism about the prospects for approaching truth over time. In the place of objectivity, they substituted an emphasis on personal and interpersonal commitments. These commitments include norms of careful investigation and accurate reporting of our findings within scientific communities, as well as furthering the broader interests of society (Polanyi, 1962, 1969). We must be willing to state our findings as claims about reality, but remain open to the possibility of revising our conclusions. Hence, there is within Polanyi's normative view both a *boldness* about stating our findings as applicable beyond the particularity of our own experiences and a *humility* that keeps us open to further learning. Our humility should lead us to continue to seek out disconfirming evidence and alternative perspectives from which to make sense of the available evidence.

Critical realism affirms both the social character of scientific knowledge and ontological realism. Critical realism allows theorizing based on unobservables, which are fundamental to the phenomena of interest in management research (Godfrey & Hill, 1995). I argued that both social construction and dynamism can be accommodated within ontological realism. The two volumes edited by Ackroyd and Fleetwood (2000; Fleetwood & Ackroyd, 2004) provide arguments for critical realism in management research and examples of its application. Critical realism is a viable alternative to positivism, postmodernism, and pragmatism as a philosophy of science for management research. Management researchers can draw upon not only the critical realist writings within philosophy of science, but also growing streams of critical realist research in the social sciences (see Archer et al., 1998).

Methodological pluralism results from both diverse traditions converging within our research field and the complexity of organizations themselves. Gadamer's work challenges us to respond to pluralism by recognizing the hermeneutical aspects of our own research (see also Bernstein, 1983). As management researchers, we face the dual hermeneutical tasks of interpreting data from organizations and interpreting others' research. The literature on hermeneutics highlights and explains the role of interpretation in the social sciences. This literature may inform our dialog among ourselves, with managers, and across disciplines. The literature on hermeneutics challenges us to reflect carefully upon how we read and write research within our eclectic field. Huff (1999) reminded us that our scholarly writing is part of a conversation, and she offered many practical suggestions to enhance our

communication. We should seek to make our research accessible to readers who may be interested in our findings but unfamiliar with our methods.

The pluralistic context of management research methods presents a daunting array of possibilities and challenges. Our human limitations force us to make trade-offs between depth and breadth in developing methods competencies. The training of researchers must include opportunities to develop along both dimensions in order to (1) read and review literature in our field and (2) make a contribution within an area of specialization. Scholars take a variety of approaches to the breadth-depth trade-off over the course of their careers. Specialists contribute by demonstrating cutting-edge applications. Generalists contribute by drawing upon a broad range of prior research findings. Both specialists and generalists play vital roles in the review process and, more generally, in advancing theory.

NOTES

1. Hermeneutics is the branch of philosophy that addresses the interpretation of texts. We can see the relevance of hermeneutics to the social sciences by recognizing that human actions, like texts, require interpretation of their meanings (Ricoeur, 1981, Chap. 8; Taylor, 1971).
2. Throughout this chapter, the term "management research" refers to research in strategic management and organization theory.
3. *Truth and Method* was originally published in 1960 as *Wahrheit und Methode* (Tübingen, Germany: J.C.B. Mohr). The 2002 revised English-language edition is based on the revised and expanded 5th German edition of *Gesammelte Werke*, Vol. 1, published in 1986 (Tübingen, Germany: J.C.B. Mohr).
4. To be consistent with Gadamer, I use the singular "method" as a broad category referring to the method of natural sciences. Later, in the discussion of management research, I introduce the plural "methods" to reflect the variety of data gathering and analytical techniques used in our field (see Hitt et al., 1998).
5. In this section, page numbers without a specific reference refer to Gadamer (2002).
6. In Gadamer's usage, "problem" refers to a dilemma for which there are no clear arguments leading to a resolution (see Gadamer, 2002, pp. 376–377).
7. Ontology addresses the question of existence and the nature of being.
8. Karl Popper expressed agreement with this perspective (see Grondin, 2003, pp. 298–299).
9. Originally published in 1970 in *Zur Logik der Sozialwissenschaften* (Frankfurt am Main, Germany: Suhrkamp Verlag, pp. 251–290).
10. In this section, page numbers without a specific reference refer to the corrected edition of *Personal Knowledge* (Polanyi, 1962).
11. Simon (1973) took the alternative view that scientific discovery could be explained as pattern induction following codifiable heuristics.
12. For more on what makes research interesting, see Davis (1971).

13. References here are to the second edition (Bhaskar, 1978).
14. For the subsequent round of responses, see Pfeffer (1995) and Van Maanen (1995a).

ACKNOWLEDGMENTS

My thanks go to Frances Fabian, Eric Tsang, participants in the Strategic Management Workshop at Purdue University, and the editors of this volume for helpful comments on earlier drafts.

REFERENCES

Ackroyd, S., & Fleetwood, S. (2000). Realism in contemporary organization and management studies. In: S. Ackroyd & S. Fleetwood (Eds), *Realist perspectives on management and organizations* (pp. 3–25). New York: Routledge.

Archer, M. S. (1995). *Realist social theory: The morphogenetic approach*. Cambridge, U.K.: Cambridge University Press.

Archer, M. S. (1996). *Culture and agency: The place of culture in social theory* (revised ed.). Cambridge, U.K.: Cambridge University Press.

Archer, M., Bhaskar, R., Collier, A., Lawson, T., & Norrie, A. (Eds) (1998). *Critical realism: Essential readings*. London, U.K.: Routledge.

Argyris, C. (1968). Some unintended consequences of rigorous research. *Psychological Bulletin*, *70*, 185–197.

Astley, W. G. (1985). Administrative science as socially constructed truth. *Administrative Science Quarterly*, *30*, 497–513.

Ayer, A. J. (1952). *Language, truth and logic* (2nd ed.). New York: Dover Publications.

Barley, S. R. (1990). Images of imaging: Notes on doing longitudinal field work. *Organization Science*, *1*, 220–247.

Berger, P. L., & Luckmann, T. (1967). *The social construction of reality: A treatise in the sociology of knowledge*. New York: Random House.

Bernstein, R. J. (1983). *Beyond objectivism and relativism: Science, hermeneutics, and praxis*. Philadelphia, PA: University of Pennsylvania.

Bettis, R. A. (1991). Strategic management and the straightjacket: An editorial essay. *Organization Science*, *2*, 315–319.

Bhaskar, R. (1978). *A realist theory of science* (2nd ed.). Hassocks, U.K.: Harvester Press.

Bonjour, L. (1978). Can empirical knowledge have a foundation? *American Philosophical Quarterly*, *15*, 1–13.

Bradbury, H., & Lichtenstein, B. M. B. (2000). Relationality in organizational research: Exploring the space between. *Organization Science*, *11*, 551–564.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Education Researcher*, *18*(1), 32–42.

Brown, J. S., & Duguid, P. (1991). Organizational learning and communities-of-practice: Toward a unified view of working, learning, and innovation. *Organization Science*, *2*, 40–57.

Brown, J. S., & Duguid, P. (2001). Knowledge and organization: A social-practice perspective. *Organization Science*, *12*, 198–213.

Canella, A. A., Jr., & Paetzold, R. L. (1994). Pfeffer's barriers to the advance of organizational science: A rejoinder. *Academy of Management Review*, *19*, 331–341.

Carnap, R. (1974). *An introduction to the philosophy of science*. New York: Basic Books.

Chia, R. (1997). Essai: Thirty years on: From organizational structures to the organization of thought. *Organization Studies*, *18*, 685–707.

Cohendet, P., & Llerena, P. (2003). Routines and incentives: The role of communities in the firm. *Industrial and Corporate Change*, *12*, 271–297.

Daft, R. L., & Lewin, A. Y. (1990). Can organization studies begin to break out of the normal science straitjacket? An editorial essay. *Organization Science*, *1*, 1–9.

Davis, M. S. (1971). That's interesting!: Towards a phenomenology of sociology and a sociology of phenomenology. *Philosophy of the Social Sciences*, *1*, 309–344.

Fabian, F. H. (2000). Keeping the tension; Pressures to keep the controversy in the management discipline. *Academy of Management Review*, *25*, 35–371.

Feyerabend, P. (1975). *Against method: Outline of an anarchistic theory of knowledge*. London: NLB.

Fleetwood, S., & Ackroyd, S. (2004). *Critical realist applications in organisation and management studies*. London, U.K.: Routledge.

Foss, N. J. (2003). Bounded rationality and tacit knowledge in the organizational capabilities approach: An assessment and a re-evaluation. *Industrial and Corporate Change*, *12*, 185–201.

Gadamer, H.-G. (2002). *Truth and Method*, (2nd, revised ed.) (Translation revised by J. Weinsheimer and D. G. Marshall) New York: Continuum Publishing.

Gioia, D. A., & Pitre, E. (1990). Multiparadigm perspectives on theory building. *Academy of Management Review*, *15*, 584–602.

Giddens, A. (1979). *Central problems in social theory: Action, structure and contradiction in social analysis*. Berkeley, CA: University of California Press.

Godfrey, P. C., & Hill, C. W. L. (1995). The problem of unobservables in strategic management research. *Strategic Management Journal*, *16*, 519–533.

Greenwood, D. J., & Levin, M. (1998). *Introduction to action research: Social research for social change*. Thousand Oaks, CA: Sage.

Grondin, J. (2003). *Hans-Georg Gadamer: A biography*. New Haven, CT: Yale University Press.

Habermas, J. (1977). A review of Gadamer's *Truth and method*. In: F. R. Dallmayr & T. A. McCarthy (Eds), *Understanding and social inquiry* (pp. 335–363). Notre Dame, IN: University of Notre Dame Press.

Habermas, J. (1984). *The theory of communicative action*. Boston, MA: Beacon Press.

Hambrick, D. C. (1994). What if the Academy actually mattered? *Academy of Management Review*, *19*, 11–16.

Harré, R. (1970). *The principles of scientific thinking*. Chicago, IL: University of Chicago Press.

Heidegger, M. (1962). *Being and time*. New York: Harper & Row.

Hendry, J. (2000). Strategic decision making, discourse, and strategy as social practice. *Journal of Management Studies*, *37*, 955–977.

Heracleous, L. (2003). *Strategy and organization: Realizing strategic management*. Cambridge, U.K.: Cambridge University Press.

Hitt, M. A., Gimeno, J., & Hoskisson, R. E. (1998). Current and future research methods in strategic management. *Organizational Research Methods*, *1*, 6–44.

Huff, A. S. (1999). *Writing for scholarly publication*. Thousand Oaks, CA: Sage.
Jarzabkowski, P. (2003). Strategic practices: An activity theory perspective on continuity and change. *Journal of Management Studies*, *40*, 23–55.
Johnson, G., Melin, L., & Whittington, R. (2003). Micro strategy and strategizing: Towards an activity-based view. *Journal of Management Studies*, *40*, 3–22.
Knorr Cetina, K. (1999). *Epistemic cultures: How the sciences make knowledge*. Cambridge, MA: Harvard University.
Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago, IL: University of Chicago Press.
Kwan, K., & Tsang, E. W. K. (2001). Realism and constructivism in strategy research: A critical realist response to Mir and Watson. *Strategic Management Journal*, *22*, 1163–1168.
Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In: I. Lakatos & A. Musgrave (Eds), *Criticism and the growth of knowledge* (pp. 91–196). Cambridge: Cambridge University.
Laudan, L. (1977). *Progress and its problems: Toward a theory of scientific growth*. Berkeley, CA: University of California.
Laudan, L. (1996). *Beyond positivism and relativism: Theory, method, and evidence*. Boulder, CO: Westview Press.
Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, U.K.: Cambridge University Press.
Lee, A. S. (1991). Integrating positivist and interpretive approaches to organizational research. *Organization Science*, *2*, 342–365.
Lee, G. K., & Cole, R. E. (2003). From a firm-based to a community-based model of knowledge creation: The case of the Linux kernel development. *Organization Science*, *14*, 633–649.
Lewis, M., & Grimes, A. J. (1999). Metatriangulation: Building theory from multiple paradigms. *Academy of Management Review*, *24*, 672–690.
MacIntyre, A. (1984). *After virtue: A study of moral theory* (2nd ed.). Notre Dame, IN: University of Notre Dame Press.
Mahoney, J. T. (1993). Strategic management and determinism: Sustaining the conversation. *Journal of Management Studies*, *30*, 173–191.
March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science*, *2*, 71–87.
McKelvey, W. (1997). Quasi-natural organization science. *Organization Science*, *8*, 352–380.
Mintzberg, H. (1990). Strategy formation: Schools of thought. In: J. W. Fredrickson (Ed.), *Perspectives on strategic management* (pp. 105–235). New York: Harper & Row.
Mir, R., & Watson, A. (2000). Strategic management and the philosophy of science: The case for a constructivist methodology. *Strategic Management Journal*, *21*, 941–953.
Mir, R., & Watson, A. (2001). Critical realism and constructivism in strategy research: Toward a synthesis. *Strategic Management Journal*, *22*, 1169–1173.
Montgomery, C. A., Wernerfelt, B., & Balakrishnan, S. (1989). Strategy content and the research process: A critique and commentary. *Strategic Management Journal*, *10*, 189–197.
Montgomery, C. A., Wernerfelt, B., & Balakrishnan, S. (1991). Strategy and the research process: A reply. *Strategic Management Journal*, *12*, 83–84.
Perrow, C. (1994). Pfeffer slips!. *Academy of Management Review*, *19*, 191–194.
Pettigrew, A. M. (1990). Longitudinal field research on change: Theory and practice. *Organization Science*, *1*, 267–292.

Pfeffer, J. (1993). Barriers to the advance of organizational science: Paradigm development as a dependent variable. *Academy of Management Review*, *18*, 599–620.

Pfeffer, J. (1995). Mortality, reproducibility, and the persistence of styles of theory. *Organization Science*, *6*, 681–686.

Pickering, A. (Ed.) (1992). *Science as practice and culture*. Chicago, IL: University of Chicago Press.

Polanyi, M. (1946). *Science, faith and society*. Chicago, IL: University of Chicago Press.

Polanyi, M. (1962). *Personal knowledge: Toward a post-critical philosophy*. Chicago, IL: University of Chicago Press.

Polanyi, M. (1969). *Knowing and being: Essays by Michael Polanyi*. London: Routledge & Kegan Paul.

Rescher, N. (1988). *Rationality: A philosophical inquiry into the nature and the rationale of reason*. Oxford: Clarendon Press.

Ricoeur, P. (1981). *Hermeneutics and the human sciences*. Cambridge, U.K.: Cambridge University Press.

Rouse, J. (1996). *Engaging science: How to understand its practices philosophically*. Ithaca, New York: Cornell University.

Sandelands, L., & Drazin, R. (1989). On the language of organization theory. *Organization Studies*, *10*, 457–478.

Scherer, A. G. (1998). Pluralism and incommensurability in strategic management and organization theory: A problem in search of a solution. *Organization*, *5*, 147–168.

Schrag, C. O. (1986). *Communicative praxis and the space of subjectivity*. Bloomington, IN: Indiana University.

Schrag, C. O. (1992). *The resources of rationality: A response to the postmodern challenge*. Bloomington, IN: Indiana University Press.

Scott, D. (1985). *Everyman revived: The common sense of Michael Polanyi*. Grand Rapids, MI: William B. Eerdmans.

Seth, A., & Zinkhan, G. (1991). Strategy and the research process: A comment. *Strategic Management Journal*, *12*, 75–82.

Shults, F. L. (2003). *Reforming theological anthropology: After the turn toward relationality*. Grand Rapids, MI: William B. Eerdmans.

Simon, H. A. (1973). Does scientific discovery have a logic? *Philosophy of Science*, *40*, 471–480.

Simon, H. A. (1991). *Models of my life*. New York: Basic Books.

Starbuck, W. H. (2004). Why I stopped trying to understand the real world. *Organization Studies*, *25*, 1233–1254.

Stenmark, M. (1995). *Rationality in science, religion, and everyday life: A critical evaluation of four models of rationality*. Notre Dame, IN: University of Notre Dame Press.

Strauss, A. (1987). *Qualitative analysis for social scientists*. Cambridge, U.K.: Cambridge University Press.

Taylor, C. (1971). Interpretation and the sciences of man. *The Review of Metaphysics*, *25*, 3–51.

Tsoukas, H. (2003). Do we really understand tacit knowledge? In: M. Easterby-Smith & M. A. Lyles (Eds), *The Blackwell handbook of organizational learning and knowledge management* (pp. 410–427). Oxford, U.K.: Blackwell.

Tsoukas, H., & Chia, R. (2002). On organizational becoming: Rethinking organizational change. *Organization Science*, *13*, 567–582.

Tsoukas, H., & Knudsen, C. (2002). The conduct of strategy research. In: A. Pettigrew, H. Thomas & R. Whittington (Eds), *Handbook of strategy and management* (pp. 411–435). London, U.K.: Sage Publications.

van Huyssteen, J. W. (1999). *The shaping of rationality: Toward interdisciplinarity in theology and science*. Grand Rapids, MI: William B. Eerdmans.

Van Maanen, J. (1995a). Fear and loathing in organization studies. *Organization Science*, *6*, 687–692.

Van Maanen, J. (1995b). Style as theory. *Organization Science*, *6*, 133–143.

Warnke, G. (1987). *Gadamer: Hermeneutics, tradition and reason*. Stanford, CA: Stanford University Press.

Webb, E., & Weick, K. E. (1979). Unobtrusive measures in organizational theory: A reminder. *Administrative Science Quarterly*, *24*, 650–659.

Weick, K. E. (1979). *The social psychology of organizing* (2nd ed.). New York: Random House.

Weick, K. E. (1989). Theory construction as disciplined imagination. *Academy of Management Review*, *14*, 516–531.

Whittington, R. (1996). Strategy as practice. *Long Range Planning*, *29*(5), 731–735.

Whittington, R. (2002). Practice perspectives on strategy: Unifying and developing a field. *Academy of Management Proceedings*, Denver, CO.

CHALLENGES AND GUIDELINES FOR CONDUCTING INTERNET-BASED SURVEYS IN STRATEGIC MANAGEMENT RESEARCH

Zeki Simsek, John F. Veiga and Michael H. Lubatkin

ABSTRACT

Given the ubiquity of Internet access in the business world, the question for strategy researchers is no longer over whether or not Internet surveys are viable, but rather over the comparative advantages and disadvantages of this modality. To address this question, we provide guidelines for researchers to help minimize the challenges while still reaping the benefits. We begin by first defining Internet survey modalities and some of their benefits, and then we focus on the associated sampling challenges and often ways that strategy researchers can address them. To further assist researchers in using this survey modality, we present a comparison of some software packages that might be useful, followed by a discussion of the lessons that we have learned from our own use of Internet surveys.

Research Methodology in Strategy and Management
Research Methodology in Strategy and Management, Volume 2, 179–196

ISSN: 1479-8387/doi:10.1016/S1479-8387(05)02008-4

INTRODUCTION

The axiom "If you want to know what managers are thinking, you have got to ask them," still holds true today, so strategic management researchers are always looking for new and better ways to query managers. Until recently, mail questionnaires, field interviews, and telephone surveys were the only convenient techniques to collect survey information. However, the emergence of the Internet, and the computer-mediated self-administered communication medium that it offers through email and web-based surveys, is changing all this.

Compared to other survey techniques, Internet surveys offer the promise of faster data gathering, error-free data entry, and lower cost. In addition, because the Internet obliterates time zones and geographic borders, surveying with it can prove very beneficial when the sample population is mobile or resides in multiple locations or countries. This technique may also be beneficial in surveying individuals who may not be willing to grant the time for personal or phone interviews, but who might respond to an email survey at their own convenience. Despite the upside, this technique also poses some challenges for the researcher, including sampling issues related to how representative the sample is, sample frame, and sampling control, and with non-sampling errors, such as non-response and measurement. These challenges, if not adequately addressed, can undermine the validity and reliability of the inferences about the researcher's focal population. Our goal in this chapter is to provide guidelines for reaping the benefits of Internet surveys by minimizing their challenges. We begin by defining Internet surveys modalities and some of their benefits, and then we focus on the associated sampling challenges followed by a discussion on how strategy researchers can address them. To assist researchers in using this survey modality, we present a comparison of some software packages that might be useful, followed by a discussion of the lessons we have learned from our own use of Internet surveys.

SURVEY MODALITIES AND BENEFITS

Fundamentally, the two Internet surveying modalities available, i.e., email and web-based surveys, differ in the manner in which respondents are identified and contacted. The email survey involves a computerized self-administered questionnaire, in which the researcher sends and the respondents receive, complete, and return it through email systems (Simsek

& Veiga, 2000). The researcher might either send an email message with the survey as part of the message text, or send an email message with the survey as an attachment that the respondent must open in order to respond. Alternatively, the researcher might embed a URL message in the email's text so that the recipient is simply directed to click on this hypertext link, which then evokes their web browser, presenting them with a web-based survey.

The web-based survey, which currently receives the most attention from researchers, involves a computerized, self-administered questionnaire in which the researcher posts the survey on a World Wide Web site where individuals access and complete the questionnaire by using compatible web browsers (Simsek & Veiga, 2001). Respondents can be diverged to the web site through links to other web pages, or invited to the web site through various means such as postal notification or email, with URL embedded links.

Both survey modalities offer the researcher cost, data-collection speed, and media richness benefits over the conventional survey techniques. The primary costs of Internet surveys include assembling and obtaining sampling frames, creating or buying software and supporting databases, and accessing the Internet. No paper is required and a direct transfer from the form to the analysis software simplifies data analysis. Furthermore, while the costs of the other techniques tend to be proportional to the size of the sample, the cost associated with adding additional respondents in Internet surveys is low to none. Any incremental cost will be limited to the cost of additional storage space allotted for returns, bandwidth load, and server capacity.

Regarding data-collection speed, Internet surveys offer the possibility of very rapid surveying, in that it can be sent as easily to a thousand people as it can to one, and all potential respondents can immediately receive the questionnaire regardless of their location. This can be especially valuable at the pilot-testing stage of survey development, where pilot testing and instrument clarification is needed before the final survey can be launched. Internet surveys also save all the time that the conventional surveys require for photocopying questionnaires, stuffing envelopes, addressing outgoing mail, and sending follow-up questionnaires.

Regarding media richness, both Internet survey modalities are media lean, in that they involve lower transmission of non-verbal cues, of varied language, of timely feedback, and of personalization, compared to other surveying techniques. On the other hand, they allow for the transmission of many different types of cues, such as text, sound, graphics, and live interaction and personal contact (e.g., via email), which adds some richness.

Media richness can also be added by using a common gateway interface script, which allows for adaptive questioning in which questions that are asked of a respondent depend upon his or her answers to previous questions. Richness can be added by adapting the Internet surveys to include timely feedback displays that are specifically tailored to the content of responses supplied by the user. For example a web-based survey can be designed so that as the survey proceeds, the questions presented are dependent upon the respondent's previous responses, a technique known as "item branching." Finally, richness can be added by adapting the Internet surveys to ensure that respondents answer all questions that are necessary before completing other aspect of the survey. As we will discuss, the added richness afforded by Internet survey modalities can be particularly important in affecting response quality and quantity.

SAMPLING AND NON-SAMPLING CHALLENGES

Researchers have noted that the method and medium in which a researcher gathers data may affect the quantity and quality of data gathered (e.g., Babbie, 1998). A number of studies have examined the advantages and disadvantages of various data-collection methods including personal interviews, telephone interviews, mail surveys, and electronic mail surveys; however, few studies have examined surveys administered on the World Wide Web.

Of the various sampling issues faced by researchers to ensure the validity and reliability of the inferences about a target population, the three that are the most challenging to Internet surveys have to do with representativeness, frame, and control. Representativeness has to do with the extent to which the sample represents the population from which it was drawn. This may prove difficult to achieve using Internet surveys for some population, particularly those where a large percentage of its members do not use the Internet, or dislike the experience of participating in electronic surveys for reasons that might be systematically tied to background, education, gender, and the like. On the other hand, representativeness becomes much less of an issue if the focal population is computer-literate and computer-willing, or when a selection bias is appropriate. For example, a researcher might want to survey opinions related to a new software program, and therefore want responses only from that subset of the population who has an informed opinion.

Sampling frame poses another challenge to Internet surveys. Few master directories exist that lists individuals (and their email addresses) from a particular population that has access to the Internet, and the few that do exist, like commercially available lists, may be seriously flawed. For example, they may include only a small percentage of all Internet users because of how they are constructed, and those on the list may have been included for reasons other than their willingness to participate in unsolicited Internet surveys, or may be listed with out-of-date email addresses. While these flaws can often be documented using traditional survey techniques like telephone and mail surveys for the purpose of determining response rates, they are virtually impossible to document using Internet surveys.

Finally, Internet surveys generally suffer from a lack of adequate sampling control. Unlike the more traditional survey techniques, it is difficult to approximate the size of the respondent pool in comparison to the size of the population and the sampling pool using Internet surveys, and thus problematic to generalize research findings beyond those responding to the survey. Also unlike the more traditional survey techniques, web-based surveys are particularly vulnerable to problems stemming from false identities. Anyone outside the target population can respond to the survey and the same individual can submit multiple responses without being detected. Indeed, stories abound about individuals misrepresenting such demographic characteristics as age, gender, level of education, and so forth. However, email based surveys are less prone to such fraud because the researcher sends the survey to individuals identified *a priori*, thereby offering greater sampling control.

In addition to the above-mentioned three forms of sampling errors, Internet survey techniques are also vulnerable to one form of non-sampling error having to do with non-response, but not the form having to do with measurement. Regarding non-response errors, a high number of non-responses raise the question of whether those who responded to the survey are different from those who did not. If non-responses are not randomly distributed – even in the absence of sampling bias – then the data generated by the survey will be biased and the inferences drawn from the data will be of uncertain validity, because non-responses compromised the assumption of sample randomization. Even randomly distributed non-responses can engender problems, if they reduce what might have been an adequate sample for hypothesis testing, in terms of size and therefore statistical power, to an inadequate one. What makes the non-response error problem particularly nettlesome with Internet surveys is that the researcher lacks an adequate sampling frame and controls to calculate non-response rates and any systematic patterns that might reside in those rates. And, without those

calculations, it is not possible to estimate whether or not the data suffer from a non-response bias.

Measurement error, the other form of non-sampling error, is represented by the deviation between the "true" and the observed responses. Broadly speaking, there are three sources of measurement error due to either the survey instrument, the data-collection technique, or the respondent. Internet-based surveys should not differ from other surveying techniques in terms of error due to the instrument, as the same steps are required to develop a reliable and valid scales. Moreover, studies that compare Internet survey techniques with traditional survey methods generally find no differences in terms of data-collection techniques. Some studies, however, find that data from Internet-based surveys are less vulnerable to respondent-based errors; i.e., the responses generated from Internet-based surveys tend to be more reliable. Researchers speculate that this might be because when the Internet gives the respondent the illusion of greater privacy, and therefore freedom to express their true thoughts and opinions without fear of being personally judged, or because respondents lose their inhibitions because of the techniques novelty (for a review of these studies, see Simsek & Veiga, 2001).

ADDRESSING SAMPLING CHALLENGES

The sampling issue about representativeness (the sample may not represent, or cover the population from which it is drawn) can be managed, in part, by using email and web-based survey modalities in combination with each other. This can allow experimentation with much more diverse population, also with population having nearly universal coverage in terms of having email addresses. It also is possible to collect survey information through an email-based survey, while posting the same survey on a bulletin board to collect information from group members who have not completed the survey.

The problem of representativeness can also be managed by using an Internet survey technique in combination with almost any other data-collection technique, including telephone interviews, personal interviews, and postal surveys. For example, because an email survey is inexpensive and fast, a researcher might begin with an email survey to determine the willingness of respondents to complete a more comprehensive Internet survey, postal survey, and so forth. Or, the researcher could simply use an email survey for those respondents who list their email addresses and use a postal survey for those without.

After the data are collected through this combination strategy, the researcher can then compare the validity of the data to those collected via traditional surveying methods. If the results are comparable then the argument could be made that the Internet sample is also representative of the general population. Moreover, when such comparable data are available, the researcher can apply post-stratification weights to the Internet data such that, for example, the number of individuals in each age, gender, and education cohort would be the same as in the population.

With regard to sampling frame challenges (directories that list individuals from a particular population), a few options exist that allow the researcher to construct a reasonable proxy frame. For example, the researcher might cull sampling frames by distributing solicitations through Listserv, discussion groups, and search engine banners. Alternatively, the researcher might cull public directories that include email addresses and are kept online by some organizations such as WhoWhere and BigFoot. Academic staff and trade association directories are, in particular, beneficial for organizational scholars because a growing number are online in publicly accessible formats. For surveys that involve employees of a single organization, we suggest first using traditional sampling frames, such as staff records, and then invite potential respondents to complete the survey. By doing so, the researcher can calculate a reasonable accurate estimate, which is perhaps the most widely refereed indicator of generalizability of survey data.

To address non-sampling errors that have to do with non-response errors and response rates, many of the techniques that have proven to be effective at improving response rates with traditional survey techniques can be adapted with minor modification to Internet survey techniques to reduce non-response errors. These techniques include monetary offerings, lottery tickets, and the possibility of winning a prize, as well as contributions to charity, offering of survey results, an appealing and/or personalized cover letter, and so forth. What researchers say to potential respondents in the survey's introductory remarks to establish legitimacy about themselves and their research project may be particularly important in affecting response rates.

Response rates can also be improved by first notifying sample members about the incoming questionnaire through an email, or postal prior-notification. Following survey convention, the prior-notification should not only seek permission, but also include: a social utility appeal that emphasizes the worthiness of the survey; an egoistic appeal that stresses the respondent's place and importance in completing the survey; and an appeal to help the researcher in completing an important project. It should also include the sponsor of the survey, a person(s) to contact for questions,

expected date of the survey, and a statement indicating the strict confidentiality of the respondent's response. When possible, the researcher should mention some possible steps that will be taken toward ensuring anonymity and confidentiality. For example, the researcher may state that screen headers will be deleted once the responses are received, offer some options for responding anonymously such as placing the questionnaire on a web site, or mention the possibility that the respondent could send the completed questionnaire through regular mail.

Response rates can also be improved by follow-up mailings. Borrowing from research about postal surveys, the underlying logic argument is that properly timed follow-up mailings provide additional stimuli for responding. And, because the cost of resending an Internet survey is trivial, follow-up mailings should include a copy of the survey as well. Response rates can also be improved by associating the study with a sponsor (an individual or institution) that is widely perceived by the focal population as being trustworthy, credible, and of high status. Response rates can be improved by offering respondents incentives as compensation for donating some of their time. This is made more possible with the advent of virtual gift-certificates that can be redeemed at web-shopping cites. And finally, questionnaire layout and design issues should be taken into account. Internet surveys in general should be accompanied by very clear and simple instructions, such as how to reply, which will neither consume much of the respondents' time nor require extensive cognitive adjustments. In particular, "extra" features that would minimize questionnaire completion time and maximize respondent convenience should be pursued, like scrolling, jump screen, quitting, no automatic next, no keyboard responses, help screens, and a progress thermometer indicating completed percentage of the questionnaire. Finally, cookies should not be used in web surveys. Many users may simply refuse to access the survey because their browsers will warn them that a cookie is being sent to their computer.

The issue of anonymity and security is particularly important in the Internet environment, in which actors have potential access to one another's personal information. The researcher should therefore take measures to alleviate respondents' concerns, like using the Usenet news groups *alt.security.pgp* and *alt.privacy.anon-server,* and directing respondents to alternative web sites like *Anonymizer.com.* There the respondent can browse the web or send email behind a firewall that the company says will render the user completely untraceable. Researchers should also set up a remailer to receive incoming electronic mail, strip the messages of the sender's identifying information and forward them anonymously to recipients – whether to

a single electronic mail box, or to thousands of addressees, or often through a series of other remailers. Because these messages are remailed in a random sequence, different from the order in which they arrive, people who may be monitoring the remailers cannot match the outgoing messages with the incoming messages to identify, who sent which message. Furthermore, the remailer itself does not store messages but serves as a channel for message transmission. That said, there is no full-proof way of guaranteeing complete data security and anonymity with Internet surveys, and many respondents know this. Thus, making the promise of complete anonymity might raise serious doubts about the credibility of the researcher.

Complete confidentiality of responses is also not possible with Internet surveys. For example, a web survey placed on an organization's web page can be accessed by anyone and even devices such as passwords and encryption may be of little use in assuring total confidentiality. Further complicating matter is the fact that many Internet users are likely to have established beliefs that Internet surveys lack confidentiality because of the popular media coverage that often calls attention to the lack of privacy on the Internet. This concern is undoubtedly higher with email surveys than they are with web-based surveys. When an email is used to send the survey to the respondent's personal email account, concerns over confidentiality such as, "if they know my email address, they also know me and how I am responding," are likely to be high. In contrast, when respondents complete a web-based survey that resides in a web page, or downloads, and returns the survey via postal services, they might be less worried about confidentiality because of greater anonymity. Regardless, any strategy taken to alleviate potential respondents' perceptions about confidentiality with Internet surveys will likely be worthwhile.

AVAILABLE SOFTWARE PACKAGES

An important aspect of the decision for using Internet surveys involves a review of the options provided by various survey software packages. There are many packages on the market, and they tend to widely differ in terms of the types of features that they offer. To help researcher evaluate these various survey software packages, we identify in Table 1 a set of criteria that a strategy researcher planning an Internet survey could use. These criteria include design capabilities, ease of implementation, ability to analyze data and/or transfer it to alternative databases like SPSS, technical resources and support, and price. For each evaluative criterion, our table indicates a set of

Table 1. Some Criteria for Internet Survey Software Evaluation.

Dimension	Evaluation Criteria
Survey design capabilities	Design surveys with no prior technical IT knowledge – no coding required Use sample questions/sample templates Self-directed survey design Self-directed survey launch Automatic response validation to keep data clean Customization of surveys for different respondents (skip logic feature) Make sure respondents complete the survey Multiple choice (dropdown list, check box) Provide open-ended questions Range type questions One-scale, two questions Force respondents to answer questions Multiple language capabilities Return to finish survey capabilities Design surveys with corporate look and feel/logo Integrated with client relationship management system Respondents management/specification of survey respondents
Technological capabilities	Software compatibility/system requirements Additional hardware requirements Externally hosted Internally hosted Technical support availability (24 h) Online support Toll-free telephone support
Survey implementation	Fast implementation Send survey invitation through emails Web link as survey invitation to survey web pages Survey pop-up in small frames within websites Control duration of survey
Survey monitoring/ analyzing/reporting	Monitor response in real-time Export data to statistical software Show basic statistical information Perform statistical analysis Basic graphic analysis (bar graphs, pie charts, histograms, etc.)
Survey price	Per-response price Price of the software Annual cost/license fee Per user

Table 1. (*Continued*)

Software Packages	Websites
SurveyView	http://www.surveyview.com/
Quask	http://www.quask.com/
Inquisite	http://www.inquisite.com/
HostedSurvey	http://www.hostedsurvey.com/
EZSurvey	http://www.raosoft.com/products/ezsurvey/index.html
Halogen	http://www.halogensoftware.com/
SurveyTracker	http://www.traintech.com/htm/software/features.htm
Websurveyor	http://www.supersurvey.com/
SurveyConnect	http://www.websurveyor.com/
FreeOnlineSurvey	http://www.surveyconnect.com/
	http://www.freeonlinesurvey.com/

issues that the researcher might consider. While our list of evaluation criteria is not exhaustive, it should help researchers evaluate different packages and decide which best fits their needs. As an aside, when selecting software, a multiple-criteria approach may prove useful. Finally, as a starting point, we provide a list of some of the most common software packages, but do not evaluate them because they are frequently updated, and each researcher needs to assess them in the context of their project.

SOME LESSONS LEARNED

Up to this point, we have addressed the benefits and challenges of Internet surveys, and have tried to provide some general suggestions for coping with these challenges. In order to provide more concrete guidance, we now share some lessons that we learned from using such surveys to collect strategy and entrepreneurship data from CEOs in small-to-medium size firms. In Fig. 1, we outline the major steps that we undertook in collecting our data, and then briefly explain what we learned with each step. We emphasize that Internet surveying, like any other mode of surveying, should first be thought of as a process.

First, we discovered that the development of a successful web-based *survey* required a considerable time investment upfront. That is, it took us more time than we had anticipated selecting a software package that best fit our research needs, programming the questionnaire, pilot testing the draft survey, and then modifying the *survey* based on results from pilot testing.

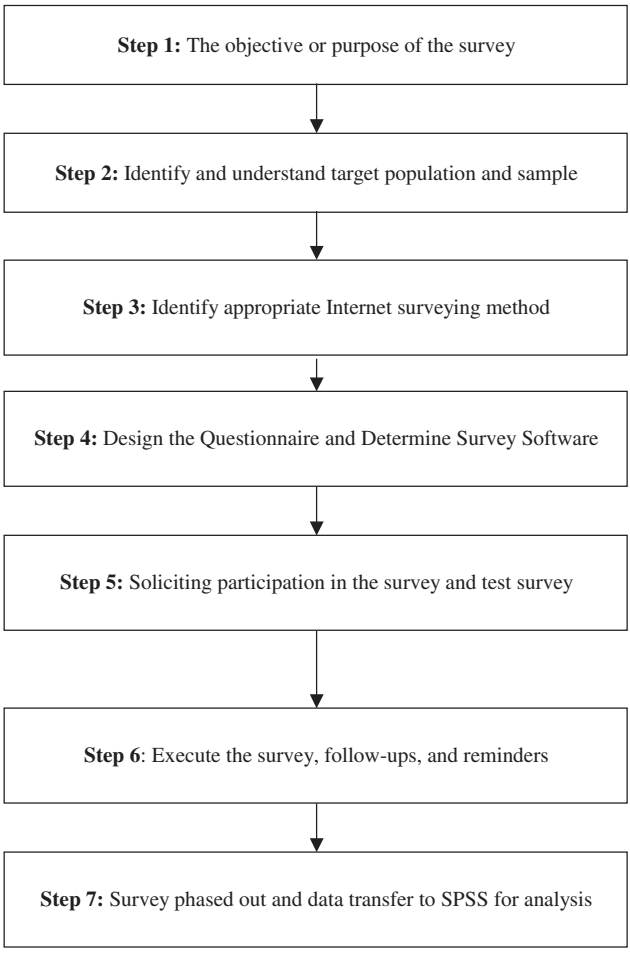

Fig. 1. Internet-based Surveying as a Process.

We began the task by clarifying the focus of our research and understanding our project's target population and sample to determine the appropriate surveying method. Then using a software package called Survey Select, we converted our standard paper-and-pencil questionnaire into an electronic form compatible with web site capabilities. Several flaws in this first effort were particularly glaring. For example, we discovered that the

survey package was not particularly flexible in generating alternative response formats such as a Likert scale or a semantic differential scale.

We also learned that long surveys are particularly not suited for the web, because the longer the survey, the longer it takes to load on a participant's computer and the greater the likelihood of non-response. Given the feedback, we received during the pilot test, we also discovered that this problem could be particularly acute, especially when respondents are using relatively slow modems. Therefore, before we conducted our final survey, we revisited our already carefully selected survey items and eliminated those that we deemed non-essential to our focal constructs. We also sought to minimize the time that respondents would have to access the survey, by dividing our survey into multiple pieces, or "web-pages." By doing this, we reduced half of the download time.

We also struggled with the decision as to whether or not we should use an identifier to link the survey to assure the validity of responses. If a web site is open to all, it is not possible to ascertain if survey respondents come from the focal population. To control "hacking" and ensure response validity, we therefore first sent emails to communicate the location of the survey to our target population.

Like any surveying situation, achieving a respectable response rate on a web survey is a challenge. Unlike a paper-and-pencil survey, you have to encourage respondents to come to your web site rather than simply responding to what is already in front of them. And, even if the population for the study is assumed to have web access, you still need to have them initiate contact with the web page in order to complete the survey. To this end, we employed different approaches to get participants to fill out our web-based evaluation forms, including email communication, using key contact people to communicate with potential respondents, an offer of survey results, and multiple follow-ups. In our particular case, we initially sent our survey, along with an emailed cover letter from the National Federation of Independent Businesses' (NFIB) president, to the CEOs of 5,957 manufacturing and service sector firms with between 20 and 500 employees. NFIB is the nation's largest small- to medium-sized (SMEs) business lobbying group with 600,000 members, and is representative of the population of SMEs in the U.S. (Dunkelberg & Scott, 1985). After three follow-ups, we received 632 responses for a response rate of 11%, "consistent with the 10–12% rate typical for mailed surveys to top executives" (Hambrick, Geletkanycz, & Fredrickson, 1993, p. 407). Moreover, we concluded that this response rate was adequate, given that our web-based survey design method, still novel to some CEOs, might have suppressed response rates below those ordinarily obtained

from mailed surveys. We excluded six firms with incomplete responses, 74 firms who reported less than 20 employees and another 10 firms that reported more than 500 employees. Our final sample consisted of 495 firms.

We also learned some "soft" lessons. The literature suggests that a "digital divide" separates those who are and are not comfortable with Internet technology and demographic variables such as race and age might be related to this problem. While we had anticipated some potential discomfort among some of the CEOs in our sampling frame, we found no evidence of a response bias in our sample, which we believe reflected a fairly high degree of sophistication among CEOs irrespective of their age or company size.

We also learned that the ease of use of distributing electronic surveys allows the researchers to efficiently deploy multiple questionnaires and frequent requests for participation, almost as an afterthought. With paper questionnaires, the barriers of printing and distribution in part hold the number of different surveys dispersed in check. With electronic surveys, third and fourth "mailings" may, thus become the expected norm although, the long-term influences of this on Internet survey response rate will need to be carefully studied and considered. In Table 2, we present an integrated set of guidelines for a strategy researcher planning an Internet survey, which are based on previously outlined steps in Fig. 1.

Although our study is encouraging, we are not suggesting that Internet surveys are always the best method for collecting strategy data. Clearly, Internet surveying makes the most sense when a large portion of the target population possesses adequate computer skills and has access to the Internet. Our project also suggests that it takes time and money to move a survey online. This phase of the process is initially more expensive and time consuming than creating a paper-based survey. The expense may not be worthwhile for single administration questionnaires that are prevalent in strategy research. Strategy researchers that frequently survey various types of surveys will probably benefit from investments in Internet survey technology; whereas those that rarely conduct surveys will see smaller returns on their investments in survey software and related technologies.

CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

Before Richard Rumelt's (1974) landmark dissertation, "Strategy, structure and performance," strategic management research was primarily case-based. With the dissertations about strategic groups by Hatten (1974)

Table 2. Some Guidelines for Designing and Implementing Internet-based Strategy Surveys.

Steps	Guidelines
The objective or purpose of the survey	Specify the population of interest Delineate the type of data to be collected Determine the desired precision of the results Consider alternative data-collection methods
Identify and understand target population and sample	Consider the viability and feasibility of Internet survey Specify the method of sample selection Consider the likely reaction of target population to alternative surveying modalities
Identify appropriate Internet surveying method	Ensure representativeness Consider potential sample size Evaluate a single or combination of Internet-surveying modalities
Design the questionnaire and determine survey software	Consider alternative survey formats Screen survey questions before using Pretest and revise the survey instrument
Soliciting participation in the survey and test survey	Prenotify that the survey has phased in Consider the most effective invitation format and method
Execute the survey, follow-ups, and reminders	Evaluate alternative response inducement techniques Consider mode and method of follow-up Use post-delivery reminder and thank-you
Survey phased out and data transfer to SPSS for analysis	Assess the survey software data transferring capability Assess the quality and quantity of the transferred data Ensure variables are being transferred correctly

and Patton (1976), and about competitive advantage by Porter (1980), the field took on a more economic bent, based primarily on large bases of secondary data like that were readily available on the Compustat data base. The trend toward the use of secondary data accelerated in the 1980s and 1990s, with others, Hambrick, who helped to introduce the PIMS database to the field in the early 1980s, Lubatkin (1983), who introduced stock

market pricing data and the CRSP data tapes. And while the field continues to rely on secondary data, more studies are recognizing the value of primary data and survey techniques to gather that data.

Indeed, data collected using surveys has contributed much to our field's recent developments. With few exceptions, most of this research has been conducted using self-administered, paper-and-pencil questionnaires distributed by mail. In this chapter, we reasoned why Internet self-administered surveys offer strategy scholars not only exciting new possibilities for data collection, but also challenges. Specifically, we highlighted challenges associated with using email and web-based surveying technique. We then suggested guidelines to help researchers improve the validity and reliability of the inferences that they draw from data about a focal population that was drawn from either of these two survey modalities. Finally, we presented several suggestions and guidelines, based on experience from one of our own projects, for strategy researchers who may be interested in collecting data using Internet surveys.

Some methodological issues about Internet surveys remain unresolved. For example, there is not much theoretical guidance to explain why some people participate as respondents in an Internet survey while others do not. Many empirical studies have been conducted on response-inducement techniques and other methodological artifacts that affect response to surveys, like preliminary notification, the foot-in-the-door technique, follow-ups, questionnaire format and length, survey source or sponsorship, nature of return envelopes, type of postage, personalization, cover letters, anonymity, deadline date, and premiums and rewards. However, there are no such studies that we are aware of that do the same for Internet surveys. There is also no clear guidance for better ensuring accurate and truthful reporting of data. Clearly, the slow development of a generally accepted theory about Internet survey design hinders the full potential of Internet surveys. However, a number of theories have been suggested as being applicable to the survey response decisions in general (Dillman, 1978; Yammarino, Skinner, & Childers, 1991). The three most cited theories that are applicable to marketing research are exchange (Dillman, 1978), cognitive dissonance (Furse & Stewart, 1984), and self-perception (Tybout & Yakch, 1980). In addition, Albaum (1987) has suggested that a theory of commitment or involvement might also fill the gap in theory development and use by marketing researchers. Clearly, future research is needed that builds and examines conceptual models based on these theories to generate a more complete understanding of why some people respond, whereas others do not, and among respondents, why the quality of data varies.

With these methodological issues in mind, we nevertheless believe that email and web-based surveys offer a promising new and efficient means for strategy researchers to query managers, and this should only improve with time. As with any evolving data-collection tool, however, the use of various Internet survey modalities must be carefully examined and should be seen as an additional supporting methodology rather than as an alternative to traditional data collection approaches. Wherever Internet surveys are used, sample representativeness and response rates are critical in evaluating the value of Internet surveys for strategy research. While such issues may become less of a problem in the future, the challenge for strategy researchers is to conduct their own research on the coverage, nonresponse, and measurement error properties of the various Internet-based survey modalities. Only by fully understanding, both the benefits and the drawbacks can the strategy researchers fully exploit the potential of web surveys. Nevertheless, these modalities offer exciting new opportunities for data collection, which should serve strategy research well in the future.

REFERENCES

Albaum, G. (1987). Do source and anonymity affect mail survey results? *Journal of the Academy of Marketing Science*, *15*, 74–81.

Babbie, E. (1998). *The practice of social research*. Belmont, CA: Wadsworth Publishing.

Dillman, D. (1978). *Mail and telephone surveys: The total design method*. New York: Wiley-Interscience.

Dunkelberg, W. C., & Scott, J. (1985). *A report on the representativeness of the National Federation of Independent Business Sample of Small firms in the U.S.* Washington, DC: Office of Advocacy, U.S. Small Business Administration.

Furse, D. H., & Stewart, D. (1984). Manipulating dissonance to improve mail survey response. *Psychology and Marketing*, *1*, 79–94.

Hambrick, D., Geletkanycz, M., & Fredrickson, J. (1993). Top executive commitment to the status quo: Some tests of its determinants. *Strategic Management Journal*, *14*, 401–418.

Hatten, K. (1974). *Strategic models in the brewing industry*. Ph.D. dissertation, Purdue University.

Lubatkin, M. (1983). Mergers and the performance of the acquiring firm. *Academy of Management Journal*, *8*(2), 218–225.

Patton, G. (1976). *A simultaneous equation model of corporate strategy: The case of U.S. Brewing*. Ph.D. dissertation, Purdue University.

Porter, M. (1980). *Competitive strategy: Techniques for analyzing industries and competitors*. New York: Free Press.

Rumelt, R. (1974). *Strategy, structure, and economic performance*. Boston, MA: Harvard Business School.

Simsek, Z., & Veiga, J. (2000). The electronic survey technique: An integration and assessment. *Organizational Research Methods*, *3*, 92–114.

Simsek, Z., & Veiga, J. F. (2001). A primer on Internet self-administered surveys. *Organizational Research Methods Journal*, *4*, 218–235.

Tybout, A., & Yakch, R. (1980). The effect of experience: A matter of salience? *Journal of Consumer Research*, *6*, 406–413.

Yammarino, F. J., Skinner, S. J., & Childers, T. L. (1991). Understanding mail survey response behavior: A meta-analysis. *Public Opinion Quarterly*, *55*, 613–639.

MULTI-THEORETICAL MIXED-LEVEL RESEARCH IN STRATEGIC MANAGEMENT

Caron H. St. John

ABSTRACT

The purpose of this chapter is to offer a discussion of the key issues in mixed-level, multi-theoretical research in strategic management. Mixed-level issues are grouped into two categories: (1) measurement of constructs, with discussion of situations in which the level of theory, level of measurement, and level of analysis differ; and (2) relationships among constructs, including cross-level and multilevel models. Key theories and views found in the strategic management literature are discussed briefly to illustrate the basic arguments of each, its focal unit of analysis, and the implicit or explicit incorporation of mixed-level perspectives.

> Organizational researchers will never be better than psychologists at understanding individuals in general, better than economists at studying large-scale market forces, nor better than sociologists at studying social forces. Only an organizational science can address effectively the complexities of the relationships between the units of different levels of analysis that comprise organizations (House, Rousseau, & Thomas-Hunt, 1995).

Research Methodology in Strategy and Management
Research Methodology in Strategy and Management, Volume 2, 197–223

ISSN: 1479-8387/doi:10.1016/S1479-8387(05)02009-6

In the 1979 edited volume of *Strategic Management: A New View of Business Policy and Planning*, the editors, Dan Schendel and Charles Hofer (1979, p. 391), provided a prescient overview of the challenge strategic management researchers would face as the field moved toward a more scientifically based, empirically oriented research discipline. In calling for more theory building and theory testing, they referenced the continuum between "the 'universal truths' type of theories and the assumption that 'each situation is unique' " (p. 391). They pointed out the opportunity to build global theories from the existing limited-domain and contingency theories by first understanding the boundary conditions of the theories and the level of the organization to which the theories apply. Twenty-five years later, it is evident that many strategic management researchers have risen to the call. Mixed-level research is commonplace and multi-theoretical research is growing in importance every year.

As described by Oliver Williamson (1999), "Business strategy is an expansive enterprise… strategy is, by nature, interdisciplinary. All of the social sciences – especially economics and organization theory… are implicated (p. 1087)." Strategic management research is concerned with individual decision makers, management teams, departments, divisions, firms, populations of like-firms, and whole industries and organization fields. These many different perspectives require theoretical contributions from a range of social science disciplines. Furthermore, unlike other research disciplines that may successfully isolate their units of analysis from contextual influences, strategy researchers often must consider their units of analysis within a larger context: individuals within groups, groups within firms, and firms within industries.

The challenges strategic management researchers face, and the value they provide, are similar to that of researchers in the medical sciences and many engineering disciplines. In discussing the organization sciences, House et al. (1995) noted, "several scientific disciplines recognize that components and context interact to create distinctive wholes. In biology, for example, cross level interactions are an essential subject of study (p. 85)." In the past, atoms, molecules, cells, genes, systems (e.g., circulatory), processes (e.g., enzymatic), organs, and humans were the domain of different scientific disciplines: chemistry, biochemistry, microbiology, genetics, various medical specialties, and practicing clinicians. The more scientists discover about each level in the complex human system, the more compelling are the questions about interconnectedness and interdependency among subsystems and levels, and the more blurred are the lines that divide disciplines. Similarly, in engineering, the boundaries between traditional theoretical disciplines are falling away. Many sophisticated design problems require deep understanding of the overall interconnections among mechanical systems, electronic systems, advanced

materials, and other changing fields – with some of the most interesting opportunities for innovation and creativity at the intersection of these fields.

For strategic management researchers, the complex system is the organization, which itself is a collection of individuals who are grouped into interdependent subsystems and levels, motivated to purposeful action, and existing within a larger, influential environmental context. As noted by many others, organizations do not make decisions, people do; organizations do not behave, people do. It is the collective action of individuals, accumulated over time, and influenced by context and process that determines the capabilities and strategies of firms, and the relative value in the marketplace of those accumulated choices. The complex system calls for continued, highly integrated contributions from behavioral, economic, and social science theories and deeper understanding of the relationships between levels of analysis.

Even though there are logical arguments for interdisciplinary, multi-theoretical work that spans levels, there are challenges as well. It is unusual for any one researcher to show mastery of the theories from so many disparate disciplines (Klein, Tosi, & Canella, 1999). Consider, for example, an area of general interest for strategic management researchers: Why do firms differ? Framed generically, which of the independent variables explain variation in the dependent variable of firm performance? To approach this research question from an integrated multi-theoretical perspective, the mass of relevant literature can be extraordinary – ranging across levels from individual managers (e.g., CEO), to management teams (e.g., top-management teams, boards of directors), to work groups (e.g., departments, divisions), to firms, to the competitive arena (e.g., geographical or strategic groups), to the larger market context (e.g., industry, including suppliers, customers, etc.), to the even larger sociopolitical context (e.g., economic and institutional forces).

The full breadth of literature underlying the theoretical streams that explain these possible independent variables and levels of analysis is unlikely to align with the education and interests of individual researchers (Klein et al., 1999). Through the influence of their major professors, many researchers align with a particular theoretical discipline in their graduate programs, form a particular mental model about appropriate theories and methods, and continue to view the world through that lens as they build their research careers. For example, a researcher aligned with the resource-based view (RBV) may choose to apply this literature stream to a series of problems or contexts. Rather than identifying complex or unresolved problems or phenomena, and searching for the one or combination of theories that provide the most effective explanation, researchers may see only those problems that fit their preferred theoretical

lens. Finally, some researchers may feel it imprudent to cast the theoretical net too widely, for fear reviewers would view their work as too broad and suggestive of trivial results (Klein et al., 1999).

In addition to the theory preferences of researchers, there are practical impediments to mixed-level, multi-theoretical research as well. The data collection efforts can be complex, time consuming, and costly (Klein et al., 1999), requiring observations at the level of individuals, firms, and industries and care taken in measurement and aggregation. Many of the difficulties are derived from misalignment among the theoretical level of constructs and the measurement complications that arise with levels, topics that will be discussed in this chapter. Misaggregation problems may be introduced when lower level measures are used to represent higher-level constructs, when higher-level measures are used to infer lower-level constructs and relationships, and when informants at the lower level are asked to report on higher-level phenomena without having the knowledge or perspective to make an accurate observation (Rousseau, 1985). Misspecification can arise when an observed relationship is attributed to a level other than the actual responding group or individual (Rousseau, 1985).

Nonetheless, the value-added that we provide as strategic management researchers is in our ability to integrate among levels and synthesize theories from different disciplines. Mixed-level theory building allows the spanning of levels to include combinations of individuals, dyads, teams, businesses, corporations, and industries, which more effectively captures the true complexity of the strategic management arena. And, importantly, it can encourage a higher level of synthesis and integration needed to explain complex problems. By explicitly considering levels issues when developing a research agenda, researchers may posit specific research questions related to changes in levels and changes in assumptions related to levels.

The purpose of this chapter is to offer a discussion of the key issues in mixed-level research. The first section provides a discussion of the construct and measurement issues associated with mixed-level research and ways of modeling causal relationships among levels. The last section offers a brief discussion of different theory streams that have been employed in strategic management research over the last 20 years and the implications of those theories for the level of analysis. As will be shown, strategic management researchers regularly employ mixed-level concepts in construct measurement, proposition development, and hypothesis testing. The methodological considerations are rarely addressed specifically, however, which means researchers are unable to accumulate deep knowledge over time with respect to these important issues.

THE CONCEPT OF LEVEL

By definition, organizations are multilevel, with individuals working in groups, groups working within organizations, and organizations operating within a larger industry environment and an even larger socio-political context. A discussion of levels often starts by specifically defining members and units, and "elements that are nested in, or members of, higher level entities." (Klein, Dansereau, & Hall, 1994, p. 198). In strategic management, examples include managers within the top-management team, members of a transaction dyad in an acquisition or joint venture, business units within a corporation, competitors within an industry, partners within an alliance, or the suppliers, buyers, and competitors that define an organizational field.

When discussing the relationship between levels, it is important to first frame the discussion into two categories (as shown in Table 1): (1) *measurement of constructs*, when data collection is conducted at different levels, and (2) specification of *relationships among constructs* that exist at different levels. The first category relates to the validity of construct measurement and is concerned with those situations where there is one construct of interest, but the theoretical level of the construct is different from the level at which measurements are taken. The second category is concerned with the specification of theoretical relationships among constructs, as when a construct at one level is related to a second construct at another level.

Construct Definition and Measurement

In mixed-level research, it is important to distinguish between the level of theory, the level of measurement, and the level of statistical analysis (Klein et al, 1994). The *level of theory* refers to the level to which generalizations will be made (Rousseau, 1985), what is typically defined as the construct of interest, the unit of analysis, or the focal point for the research. The *level of measurement* refers to the actual source of data and the *level of statistical analysis* refers to the level of the data when analysis is conducted.

It is sometimes necessary and appropriate for a researcher to define a construct at one level (the level of theory), then collect data to measure the construct at a different level (level of measurement). It is very common in strategic management research, for example, for firm-level or industry-level constructs to be measured by collecting data at the level of individuals, such as individual perceptions of industry-level environmental uncertainty,

Table 1. Possible Relationships between Levels.

Type of Mixed-level Issue	Construct Definition and Measurement	Relationships Among Constructs
	X | X	$X \rightarrow Y$ (a) $X \nearrow Y$ (b) $X \downarrow Z$, $Z \rightarrow Y$ (c) $Z \rightarrow Y$, $X \uparrow Z$ (d) $X_1 \rightarrow Y_1$ $X_2 \rightarrow Y_2$ (e)
Definition	• Level of theory and level of measurement differ	• Cross-level: Construct at one level is related to a second construct at another level (a, b, c, d) • Multi-level: Constructs and relationships between constructs at one level can be generalized to another level (e)
Concerns	• Validity of constructs, level-specific validity, direction and method of aggregation of measurements to capture constructs	• Direct (a and b above) or moderating effects (c and d above), direction of the relationship • Level-specific construct validity • Ability to generalize to more than one level

resource scarcity, and competitive intensity. If the individual data are then aggregated to form a firm-level measure of environmental uncertainty used in subsequent data analysis, then the level of statistical analysis would be the firm. For example, in a study of international joint ventures (IJV), Luo (2002) collected data from senior managers in IJVs about profitability, sales level, competitive advantage, and overall firm performance, then aggregated the measures into an overall measure of IJV performance. In this case, the level of theory is the firm, the level of data collection is the individual, and the level of statistical analysis is also the firm.

In some cases, researchers are interested in a lower-level construct, but infer that construct from higher-level measurements, following a top-down logic, as when abnormal stock market returns are used to judge the potential for synergies between two merger partners (Balakrishnan, 1988). In this case, the level of theory is the intended merged firm, but the level of measurement is higher – the market. Similarly, when R&D department capability is determined by the firm's R&D expenditures as a percentage of sales (Sakakibara, 2002), the level of theory is the R&D function, but the measurement is a financial input made at the business-unit level.

In specifying a level of theory, a researcher is implicitly or explicitly predicting that members of a group or units nested within entities are either homogeneous, independent, or heterogeneous with respect to the constructs of the theory (Klein et al., 1994), as shown in Table 2. These three sets of assumptions have significant implications for the level of measurement, the level from which data are collected, and the level of statistical analysis.

When a researcher assumes *homogeneity*, the idea is that members (individuals or groups, or organizations) are sufficiently similar on the construct of interest to be characterized as a whole. Data may be collected at the level of the member, aggregated to the level of the unit, and then used to represent a unit-level construct. An assumption of *independence* suggests

Table 2. Top-down Versus Bottom-up Methods of Data Aggregation and Forms of Relationships.

	Construct Definition and Measurement	Cross-level Relationships
Top-down	Measurements collected at a higher level used to capture a construct anchored theoretically at a lower level • Direct extrapolation • Decomposition	Higher-level factors influence lower-level phenomena • Direct effect (a, in Table 1) • Moderating effect (c, in Table 1)
Bottom-up	Measurements collected at a lower level used to capture a construct anchored theoretically at a higher level • Composition • Compilation	Lower-level factors influence upper-level phenomena • Direct effect (b, in Table 1) • Moderating effect (d, in Table 1)

that the value of the construct for a member will be independent of the value of the construct for other members of the same unit. In this case, construct, data, and measurement focus on members, independent of the unit to which the member belongs. An assumption of *heterogeneity* suggests that members vary within the unit with respect to the construct of interest, and are also influenced by their group membership. In this situation, context influences the construct of interest, data are collected from members, and member data are analyzed within the unit context.

Assumptions of homogeneity, independence, and heterogeneity also speak to the method of data aggregation, as summarized in Table 3. Mixed-level research generally follows a bottom-up or a top-down aggregation logic when the construct is at one level and measurement takes place at another level.

Bottom-up

The bottom-up approach expresses the form of relationship that exists when lower-level building blocks combine to form higher-level phenomena (Kozlowski & Klein, 2000). Bottom-up relationships between levels are of two types: composition and compilation (Kozlowski & Klein, 2000). *Composition* describes phenomena that are essentially the same as they emerge upward across levels. The phenomenon of interest at the higher level is the same, for all intents and purposes, as its component parts. An example is when organization culture is described as a composite of the beliefs and values of the individuals within the organization.

For strategy researchers, there are three basic composition models (Chan, 1998): additive, direct consensus, and dispersion (see Table 4). With the *additive* model, a higher-level construct is represented with a simple summation or an average of lower-level measurements. For *direct consensus* and *dispersion* models, the higher-level construct is tied to either consensus or dispersion among members with aggregation of within-group agreement or within-group variation scores to capture the higher-level construct.

Unlike composition models, *compilation* describes a relationship between levels in which the higher-level phenomenon is a result of a combination of related, but different, lower-level properties (Kozlowski & Klein, 2000). In other words, the construct has sub-dimensions that must be captured during the data collection process and combined appropriately to create a valid measurement. An example is organizational knowledge or knowledge inventories, which are a complex function of the acquisition, retention, deployment, idling, and abandonment of specific knowledge categories over time (Levitt & March, 1988; Levinthal & March, 1993). The Luo (2002)

Table 3. Possible Relationships between Members and Units on One Construct of Interest.

	Assumption of Homogeneity of Members within the Units	Assumption of Independence of Members from the Units	Assumption of Heterogeneity of Members within Units
Members within units. Examples include: – Individuals within groups Or – Groups within organizations Or – Organizations within industries	Members (individuals or groups or organizations) are sufficiently similar on the construct of interest to be characterized as a whole A single value of the construct is sufficient to describe the whole unit (group, organization, or industry)	Members of a unit (group, organization, or industry) are free of unit influence with respect to the construct of interest A value of the construct for a member will be independent of the value of the construct for other members of the same unit	Members vary within the unit with respect to the construct of interest Members are neither homogeneous nor independent of the group Considers the effect of context on construct of interest
Level of theory – the level to which generalizations are made	The unit (e.g., group or organization or industry)	The independent member of the unit (e.g., individual, group, organization)	The member within the context of the unit (e.g., individual within the group, the group within the organization, or the organization within the industry)
Level of measurement – the unit to which data are attached at the time of measurement	Data collected from members	Data collected from members	Data collected from members
Level of statistical analysis – the level to which data are attached at the time of analysis	Data analyzed as unit	Data analyzed as members	Data analyzed as members within unit context

Source: Adapted from Klein et al. (1994, pp. 195–229).

Table 4. Composite Models Relevant for Strategic Management.

Composition Models	Relationship between Levels	Data Handling
Additive	Higher-level is a summation of lower-level units. Variation among units is not relevant	Sum or average of lower-level scores to capture higher-level constructs
Consensus	Higher-level construct tied to consensus among members	Within-group agreement scores used to capture higher-level construct
Dispersion	Higher-level construct tied to dispersion among members	Within-group agreement scores used to capture higher-level construct

Source: Chan (1998).

study noted, earlier employed a bottom-up compilation approach. Factor loadings on the profitability, sales level, competitive advantage, and overall performance measures were used as weights in calculating average performance scores for IJV.

Top-down

Researchers may also collect data at a higher level to infer the behavior of a construct at a lower level, such as using firm-level employee turnover statistics to measure employee morale, firm-level R&D expenditures as a percent of sales to measure R&D department capability (Sakakibara, 2002), or shareholder concentration as a measure of monitoring ability in applications of agency theory (Wright, Ferris, Sarin, & Awasthi, 1996). In some instances, the higher-level measurements are chosen because of the high cost of collecting data at the level of individuals only to aggregate those data to the level of firms. When appropriate considerations of theory suggest that the higher-level measure will be a content-valid measure of the construct, then the measurement represents a type of *analytical extrapolation*. For example, the Dess and Beard (1984) measures of industry (defined by SIC or NAICS code) dynamism, complexity, and munificence are often used to describe the environment faced by a specific firm, even though there might be an argument that a firm's strategy choices have positioned it in a marginally different environmental niche or population. Even so, in this situation, the higher-level measures are directly extrapolated to the lower-level construct.

In other situations, the higher-level measurement may be *decomposed* to tease out the important lower-level construct. In a recent study by Jensen and Zajac (2004), the researchers departed from the traditional treatment of the demographic characteristics of the upper-echelon management team by disaggregating the measurement to allow consideration of the governance position, which resulted in significant, non-obvious findings. Similarly, in a study of multinational firms, each of which operated several divisions with multiple subsidiaries within each division, Frost, Birkinshaw, and Ensign (2002) argued for the importance of collecting data at the level of the subsidiary rather than just the division, because of concerns that measurement at the division level would fail to capture important aspects of inter-subsidiary coordination.

Relationships Among Constructs

The second category of mixed-level research design issues, as shown in Table 1, relates to the theoretical model used to frame propositions or hypotheses about relationships between independent and dependent variables. Roberts, Hulin, and Rousseau (1978) were among the first to forward an explicitly mixed-level conception of the organizational sciences. Rousseau (1985) developed a typology of mixed-level models that is useful for framing a discussion of alternative research designs and theoretical models. For strategic management researchers, two mixed-level research designs seem most relevant: (1) cross-level and (2) multilevel.

Suppose, for example, a researcher proposes to investigate the relationship between a construct X measured at one level and a second construct Y measured at a different level. Depending upon the research question and theoretical model, the form of the relationship might involve a *direct effect* of construct X on construct Y (examples a and b in Table 1), or a *moderating effect* of construct X on the relationship between construct Z and construct Y (examples c and d in Table 1). This type of model is referred to as *cross-level*, with independent and dependent variables measured on different levels (Rousseau, 1985). As with construct measurement, cross-level models of relationships among constructs may be characterized as top-down or bottom-up.

Cross-level, Top-down

When a cross-level research design captures a top-down relationship between levels, it establishes the role of the higher-level context on the

lower-level members. In a top-down model, the contextual factors may have a direct effect or a moderating effect. When the effect is *direct*, the contextual factors may be described as influencing the behavior of members, as when industry factors are hypothesized to influence firm performance (Caves & Porter, 1977; Hawawini, Subramanian, & Verdin, 2003), or the leadership style of a CEO influences the level of motivation or job satisfaction among employees (example a, Table 1). When the effect is *moderating*, then the contextual factors influence the relationship that exists between two or more variables at the lower level, as in the study by Carpenter (2002) that postulated strategy and social context as moderating variables in the relationship between the top management team and firm performance (example c, Table 1).

Cross-level, Bottom-up

Similarly, a cross-level research design may specify a bottom-up relationship between levels. In a bottom-up cross-level design, the independent variables are at the lower level and the dependent variables are at a higher level, as when members serve to influence, constrain, or facilitate a higher-level unit outcome. Bottom-up cross-level designs are very unusual in most organization sciences (Klein et al., 1999), but are at the core of many research questions in strategic management. The role of individuals in influencing organizational capabilities, the role of departments and functional strategies in bringing about organizational capabilities and performance, the capabilities and strategies of business units as determinants of corporate level performance, and the collective action of firms within an industry as a determinant of average industry profit – are all examples of bottom-up, cross-level research questions. For example, one of the key tenets of the RBV is that decisions made by managers (lower level) to make sustaining investments in particular resources and capabilities can bring about a firm-level competitive advantage (Barney, 1991; Dierickx & Cool, 1989). In one recent study that exemplifies a bottom-up, cross-level design, Chadwick, Hunter, and Walston (2004) proposed a relationship between the human resource management practices used by companies during layoffs and subsequent firm-level financial performance.

Multilevel Models

Multilevel models describe relationships among independent and dependent variables that can be generalized across two or more levels (Rousseau, 1985). In other words, the same relationship between independent and dependent variables at one level can be observed between analogous

independent and dependent variables at another level. House et al. (1995) offered a similar argument for mixed-level research designs, describing the approach as meso-theory building: "Meso-theory and research concerns the simultaneous study of at least two levels of analysis wherein: (1) one or more levels concern individual or group behavioral processes or variables, (2) one or more levels concern organizational processes or variables, and (3) the processes by which the levels of analysis are related are articulated in the form of bridging." (House et al., 1995, p. 83)

As a first step, a multilevel model must address constructs and components that have meaning on several levels. A study by Gersick (1991) provides an excellent example of multilevel theory building using the punctuated equilibrium model and six different levels: individuals, groups, organizations, scientific disciplines, biological systems, and general systems. She first defined each of the constructs of interest (deep structure, equilibrium periods, and revolutionary periods). After that, through an extensive literature review, she showed the commonalities in content for each construct at each level, and then built an integrated, multilevel explanation of the punctuated equilibrium theory.

There are some examples of multilevel theory building related to strategic management research that build on concepts that originating in the behavioral sciences. Following the logic of a multilevel theory, some researchers have extended the ideas of mental models and cognitive schema, developed at the level of individuals to the level of firms to form what Abrahamson and Fombrun described as a *macroculture*, the "relatively idiosyncratic organization-related beliefs that are shared among top managers across organizations (1994, p. 730)." Abrahamson and Rosenkopf (1993) took the idea of a macroculture to an even higher level in a study of interdependent networks of competitors, called collectivities, that experience close to full information about the organization processes and innovation capabilities of competitors. In a study of transformational leadership, Yammarino and Dubinsky (1994) noted that transformational leadership concepts were often extended to different levels, without proper justification and definition of boundaries. They constructed a study to specifically address individuals, the superior-subordinate dyad, and work group level issues to determine the domains to which the theory could be expected to hold.

Within mainstream strategic management research, multilevel theory building is still unusual. Although Porter (1980, 1990, 1998) extended concepts of industry structure and average industry profitability from the level of industries to the level of regional clusters and nations, it was not a specific attempt to develop a multilevel theory. Similarly, the concepts of the RBV

have been extended to additional levels. Researchers have applied RBV concepts to firms, alliance partners, supply chains, networks, and regions to explain why some groupings of firms outperform other groupings of firms (Dyer & Singh, 1998; Ettlie & Sethuraman, 2002; Park, Mezias, & Song, 2004; Wilk & Fensterseifer, 2003). As with Porter's work, these studies provide examples of efforts to use the logical arguments of the RBV to frame a study at a different level. Similarly, transaction cost economics (Williamson, 1975, 1985) and agency theories (Jensen & Meckling, 1976) have been applied to dyadic relationships at different levels, which suggests the early stages of multilevel theory building.

Thus far, these extensions have not yet risen to the level of true multilevel theory building, although they do provide some of the preliminary evidence to suggest that it may be possible to do so. Multilevel theory building would involve establishment of level-specific construct validity and clearly established relationships among constructs that can be replicated across levels. Multilevel theories continue to be unusual in the behavioral and social sciences, but provide substantial opportunities for strategic management researchers because of the inherent multilevel perspective of the field.

THEORIES, LEVELS, AND STRATEGIC MANAGEMENT RESEARCH

In 2004, Ramos-Rodriquez and Ruiz-Navarro published an extensive bibliometric analysis of the intellectual structure of strategic management research. They used citation and co-citation analysis of references listed in articles published in *Strategic Management Journal* to determine groupings of authors and topics over time. The citation analysis easily establishes industrial organization theory, represented by Porter (1980), and the RBV, represented by Barney (1991) and Wernerfelt (1984), as the two most influential literature streams in strategic management. Seminal articles associated with others of the more important research streams in strategic management are shown in Table 5. In the following section, several of the key theories and views will be defined briefly, not to provide a definite review of the literature, but to illustrate the basic arguments of each, its focal unit of analysis, and the implicit or explicit incorporation of mixed-level perspectives. The review is organized by focal unit of analysis: individuals within firms and firms within industries.

Table 5. Theory Streams in Strategic Management Research.

Literature	Level of Theory	Seminal Citations[a]
Resource-Based View	Resources within a firm	Barney (1991), Dierickx and Cool (1989), Penrose (1959), Peteraf (1993), Rumelt (1984), Wernerfelt (1984).
Strategy choice, organizational alignment, and performance	Manager within a firm, firm within an industry	Andrews (1971), Ansoff (1965), Chandler (1962), Miles and Snow (1978), Rumelt (1974), Thompson (1967)
Structure-conduct-performance branch of Industrial Organization	Firm within an industry	Porter (1980, 1981)
Transaction Cost Economics	Transactions between individuals and between firms	Williamson (1975, 1985)
Agency Theory	Individual within a firm	Jensen and Meckling (1976)
Resource Dependency	Firm within its organizational field	Pfeffer and Salancik (1978)
Organization Ecology	Population of firms	Hannan and Freeman (1984)
Institutional Theory	Organizational field, with focus on role of individuals, firms, and environments, depending upon the source of isomorphism	DiMaggio and Powell (1983)
Behavioral or Upper Echelon (Cognitive, Demographic)	Individual managers and/or board members	Hambrick and Mason (1984)

[a]"Most influential citations" as described in Ramos-Rodriquez and Ruiz-Navarro (2004).

Individuals within Firms

Research related to *upper echelons* has employed cognitive theory as well as other behavioral theories from psychology and sociology to explain the role of top executives and boards of directors within firms. These theories,

some of which have centered on the concept of mental models or cognitive schema, have been used to explore the role of cognition, experiences, and biographical factors in explaining decisions, preferences, and the effectiveness of the top executives (Hambrick & Mason, 1984; Priem, Lyon, & Dess, 1999; Wierseman & Bantel, 1992). Generally, the level of measurement is the individual, with a bottom-up composition or compilation to the level of the firm for statistical analysis. Both the level of theory and the level of statistical analysis address the board of directors or the top-management team, which, after aggregation, is treated as a firm-level variable.

The literature on *organization learning*, *absorptive capacity*, *and knowledge inventories* (i.e., the processes through which knowledge is accumulated from sources inside and outside the firm, retained and preserved, and then applied to a new situation) frames the role of human knowledge in the development of firm-level capabilities and performance (Cohen & Levinthal, 1990; Levinthal & March, 1993; Levitt & March, 1988). This view of firm-level resources builds on an assumption that the knowledge of the individuals and the abilities of individuals to learn over time from sources inside and outside the firm can be aggregated to the level of the firm following a bottom-up compilation approach to construct measurement. In testing the value of organizational knowledge as a source of competitive advantage and above-average industry profits, the causal model would be bottom-up, cross-level.

Contract or Transaction Dyads

In *agency theory*, the firm is described as a nexus of implicit and explicit contracts or agreements among participants such as owners, employees, managers, and others (Jensen & Meckling, 1976). Agency costs include losses incurred by the owner when the agent (manager) does not act in the owner's interests and it costs the owner to monitor the activities of the manager. In applications of agency theory, the level of measurement is sometimes higher than the level of the construct, which involves an extrapolation to the lower level. The *transaction cost perspective* (Williamson, 1975, 1985) has been employed to explain several strategy issues, including make-versus-buy decisions and use of alliances. In recent years, researchers have modified the transaction cost arguments to account for supply chains, networks, and alliances (Ettlie & Sethuraman, 2002; Parkhe, 1993). In this

theory, the focal unit of analysis is the transaction dyad, which usually involves transactions between firms, but within different contexts.

Firms within Industries

The Resource-Based-View was introduced into strategic management research by Barney (1991) and by Wernerfelt (1984). Within the RBV, resources and capabilities that can lead to competitive advantage are those that are unique and inimitable, from the point of view of competitors, and valuable and non-substitutable, from the point of view of customers (Barney, 1991). Firm-specific resources and complex processes and routines may serve as sources of advantage because of the causal ambiguity that is created, which, in turn, restricts the ability of competitors to imitate (Barney, 1991). The independent variables of resources and processes are defined as unique, valuable, and inimitable within a specific competitive context, defined by the competing resources and capabilities of competitors, and the needs and preferences of customers. The arguments of the RBV call for mixed-level measurements, capturing distinguishable resources within the firm and the value and uniqueness of the resources within the environmental context. The contextual aspects of competitive advantage have not been explored fully, but call for mixed-level approaches to construct definition and measurement and cross-level relationships among constructs. One of the criticisms of RBV research is the inherent tautology of postulating a relationship between a unique and valuable resource (X) and competitive advantage (Y) (Priem & Butler, 2001). A second criticism is the problem of measuring the uniqueness and value of a resource when secondary data sources are used without proper organizational and competitive context (Rouse & Daellenbach, 1999). These issues underlying construct definition, construct measurement, and contextual influences provide methodological challenges for the RBV.

Resource dependency theory takes the position that organizations may act to reduce vulnerabilities imposed by their environments and to increase their own power relative to resource holders (Pfeffer & Salancik, 1978). In resource dependency theory, organizations seek to decrease their dependence on others, and will alter their structures and patterns of behavior in order to acquire and maintain access to external resources (Pfeffer & Salancik, 1978). The focal unit of analysis is the firm, with particular emphasis on the relative power position that the firm has acquired with respect to resources of interest within the larger environment. Resource dependency theory is an interesting example of a theory that allows researchers to postulate different

directions of influence and propose different causal models to tease out interesting relationships. Resource dependency theory can be used to model top-down cross-level relationships as resource-holders exert influence over firms, or as a bottom-up cross-level relationship as firms take action to influence the power of resource holders.

The primary thesis of *institutional theory* is that organizations must conform to the established rules and norms of dominant institutions in order to gain support and be perceived as legitimate (DiMaggio & Powell, 1983). Institutional theory advances the three drivers of isomorphism or similarity that result from conformity (DiMaggio & Powell, 1983): (1) mimetic isomorphism, as firms engage in imitation to economize on search costs and cope with uncertainty, (2) coercive isomorphism, when firms comply with procedures, controls, and standards imposed by external institutions, and (3) normative isomorphism, which results from the tendency of individuals with similar backgrounds, educations, and industry experiences to define problems, filter information, and make decisions in a similar fashion (Huff, 1982; Prahalad & Bettis, 1986; Spender, 1989). Unlike some other theories that focus on explaining firm differences that give rise to performance differences, institutional theories are often employed to explain why firms are so similar. Each of the drivers of isomorphism addresses a different level of theory: mimetic isomorphism results from the imitative actions of the firm in response to competitor actions; coercive isomorphism results from the pressures imposed by institutions in the external environment; and normative isomorphism results from the similar mental models of individuals who have lived through similar life experiences or made similar decisions. As with resource dependency theory, institutional theory allows researchers to propose competing models of causality – with top-down and bottom-up models serving useful purposes in different situations and providing a tension that can lead to new research questions.

The *Bain–Mason structure–conduct–performance* (SCP) model (Porter, 1981) proposed that a firm's performance depends on the characteristics of the environment in which it competes. Industry structure, which is a function of barriers to entry, number and size of competing firms, demand elasticity, and differences among competitors, defines the conditions for industry profitability, which in turn frames the strategic options available to firms. For strategy researchers, the SCP model is usually employed to explain the role of industry characteristics on the strategy choices and performance outcomes of firms. As such, it is an example of a top-down, cross-level model with industry effects often proposed as a moderator of the strategy-performance relationship.

Multi-theoretical Research

Although there are several literature streams that have influenced the development of strategic management research, with different levels of analysis incorporated into those various perspectives, most strategic management research studies employ the concepts of one theory. As noted earlier, some strategy researchers align themselves with a particular theoretical lens – the RBV, transaction cost economics, the SCP model, or behavioral theories. Their research is theory-led, and serves to extend the usefulness of a particular theory to different contexts and, possibly, to different levels of analysis. This type of research focus provides a natural foundation for the development of true multilevel theories, which would be a very positive outcome for our field.

In recent years, several researchers have taken on the challenge of contrasting or integrating different theoretical perspectives in order to provide richer explanations of a particular phenomenon of interest. In these cases, the research is criterion-led and, in some studies, spans levels of analysis. Some studies have contrasted or integrated two separate theory streams. For example, Madhok (2002) contrasted the key questions of the RBV (why do firms differ?) with those of transaction cost theory (why do firms exist?) to illustrate the degree to which transaction cost theory anticipated the same issues raised by the RBV. In a study of the influence of mergers on firms' product-mix strategies, Krishnan, Joshi, and Krishnan (2004) employed institutional and resource-based theories to capture firm- and industry-level influences. Miller (2002) integrated the literature on knowledge management with real options theory to explain technology investments over time.

Some studies have integrated three or more theory streams to shed new light on a particular phenomena. Pouder and St. John (1996) explored the geographical clustering of competing firms and its effects on patterns of innovation over time at the level of individuals, firms, and regions by employing the RBV, cognitive theories–mental models, and institutional theories. Lee, Lee, and Rho (2002) integrated three perspectives on the emergence of strategic groups, referencing literature in cognitive and behavioral theory, industrial organization theory, and the evolutionary perspective. In an investigation of the effect of outside directors on corporate boards during institutional transactions, Peng (2004) applied agency theory, resource dependency theory, and institutional theory. Drawing on theories from economics, sociology, and the psychology of ideas, Rodan and Galvnic (2004) explored social networks, managerial performance, and innovation performance.

The above review is not definitive, but just provides some examples of the kinds of multi-theoretical work pursued by strategic management researchers. There continue to be substantial opportunities to contrast and integrate theories to offer enriched interpretations of phenomena, and to extend understanding across levels of analysis. Table 6 summarizes the different

Table 6. Framing Cross-level, Multilevel, and Multi-theoretic Research Questions.

	One Theory	Multiple Theories
Single-level	Investigate individual-level attribute as predictor of individual-level outcome Investigate group-level attribute as predictor of group-level outcome Investigate firm-level attribute as predictor of firm-level outcome Investigate industry-level attribute as predictor of industry-level outcome	Contrast ability of different theories to explain single-level relationships Integrate different theories to provide a richer explanation of a single-level relationship
Cross-level	Investigate firm-level attribute as predictor of industry-level outcome Investigate individual, group or departmental level attribute as predictor of firm-level outcome Investigate industry-level attribute as influence on firm-level outcome	Contrast ability of different theories to explain cross-level relationships Integrate different theories to provide a richer explanation (e.g., role of context) of cross-level relationships Integrate different theories to build models of relationships that consider phenomena of interest across levels
Multilevel	Investigate the explanatory power of a theory when it is applied, in total, to a new level	Contrast resilience of different multilevel theories across levels and contexts Integrate different multilevel theories to provide a holistic model of firm behavior

questions posed for single-level, cross-level, and multilevel research models, incorporating one or more theory streams.

CLOSING REMARKS

As noted by the examples provided throughout this chapter, measurements and models that span levels are common in strategic management research. Interestingly, however, the terminology used by behavioral scientists to describe these measurements and models has not been widely adopted by strategic management researchers. Without proper framing, strategic management researchers may inadvertently create the impression that we are unaware of these complex levels issues. To help communicate the thoughtfulness applied to construct measurement and models that span levels and theories, it would be helpful if we could adopt a common vocabulary and way of describing levels issues so that we accumulate knowledge and experience as a field.

Construct Measurement

Given that strategic management research so often involves collecting data at one level to measure a construct at a different level, researchers should specifically address this issue during the design, execution, and reporting of research. As researchers define constructs and measures, they should make specific reference to the level of theory, the level of measurement, and the level of statistical analysis for each. Then, if the level of theory and the level of measurement differ for a particular construct, steps should be taken to establish level-specific construct validity.

Before collecting data and aggregating measures, researchers should also specifically identify and reference the assumptions regarding homogeneity, independence, and heterogeneity with regard to members and units (see Table 3), and, for each measure, describe the method of data aggregation that will be followed: top-down (extrapolation, decomposition) or bottom-up (composition, compilation).

Model Development

Strategic management research often involves hypothesized relationships that span levels, which have been called cross-level research models throughout this chapter. By considering the relationships among levels,

strategic management researchers have the opportunity to explore issues of interaction, integration, coordination, and interdependence in ways that are not captured in single-level research (Morgeson & Hofmann, 1999) or when levels issues are not specifically addressed. In designing, executing, and reporting cross-level studies, researchers should specify the type of cross-level model (top-down or bottom-up), and any boundary conditions or underlying assumptions that frame the research model.

As noted by Rousseau (1985), there are opportunities to use the concepts above to challenge model assumptions and offer new perspectives. For example, if the assumptions about homogeneity, independence, and heterogeneity were changed, how might interpretations differ? If the level of theory or the level of measurement were changed, how would interpretations change? Are there opportunities to change the direction of hypothesized relationships among constructs – from top-down to bottom-up, or vice versa? Are there opportunities to extend constructs and relationships to new levels, thus developing multilevel models?

Contrasting and Integrating Theories

In addition to a tradition of mixed level constructs and cross-level models, strategy researchers have demonstrated a willingness to employ theories that find their roots in many disciplines: psychology, sociology, economics, and ecology. Researchers are also exploring opportunities to contrast and integrate perspectives offered by those different theories, as described earlier in this chapter. For strategic management researchers, the phenomena that provide opportunities for multilevel theory building are abundant: competitive advantage; inimitable, valuable and unique resources; innovative capacity; organizational knowledge; leadership; organization culture; and many others. These areas, and many more, require a study at the level of individuals, firms, and industry context, employing and integrating theories that address each perspective. In the coming years, it is very likely that this kind of interdisciplinary work will increase, allowing more accurate specification of the causal network acting on dependent variables of interest, which will, in turn, allow increased generalization.

Among organizational scientists, strategic management researchers are, in all probability, at the forefront of mixed-level research given the nature of the complex system and key research questions of strategists. During the last 25 years, researchers have made steady progress in incorporating new theories into the body of work and framing cross-level research models. Our

Table 7. Practicing Mixed-Level Research.

Definition and Measurement of Constructs
Define constructs and measures, making specific reference to the level of theory, the level of measurement, and the level of statistical analysis for each
If the level of theory and the level of measurement differ for a particular construct, take steps to establish level-specific construct validity
Identify and specifically reference the assumptions regarding homogeneity, independence, and heterogeneity with regard to "members" and "units"
In line with those assumptions, describe the method of data aggregation that will be followed to form measurements: top-down (extrapolation, decomposition) or bottom-up (composition, compilation)
Specification of Relationships among Constructs
Define the level for each construct in the hypothesized relationships
Describe the proposed moderating or direct effects
Specify the direction of the relationship between levels (top-down or bottom-up) – the type of cross-level model
Definition of Boundaries and Assumptions
If the assumptions about homogeneity, independence, and heterogeneity were changed, how might interpretations differ?
If the level of theory or the level of measurement were changed, how would interpretations change?
Are there opportunities to change the direction of hypothesized relationships among constructs – from single-level to cross-level? From cross-level, top-down to cross-level bottom-up, or vice versa?
Are there opportunities to extend constructs and/or relationships to new levels, thus developing multilevel models?
To provide a richer explanation, broaden boundaries, or stimulate inquiry, are there opportunities to contrast the contributions of two or more theories? To integrate two or more theories?

task, going forward, is to make our use of mixed-level approaches more explicit in our research designs and document our knowledge more systematically so that we accumulate experience with these challenging methods issues. Table 7 offers guidelines for employing mixed-level perspectives in measurement and model development.

REFERENCES

Abrahamson, E., & Rosenkopf, L. (1993). Institutional and competitive bandwagons: Using mathematical modeling as a tool to explore innovation diffusion. *Academy of Management Journal*, *18*(3), 487–518.

Andrews, K. R. (1971). *The concept of corporate strategy*. Homewood, IL: H. Dow Jones-Irwin.

Ansoff, H. I. (1965). *Corporate Strategy*. New York: McGraw-Hill.

Balakrishnan, S. (1988). The prognostics of diversifying acquisitions. *Strategic Management Journal*, *9*(2), 185–197.

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, *17*(1), 99–120.

Carpenter, M. A. (2002). The implications of strategy and social context for the relationships between top-management team heterogeneity and firm performance. *Strategic Management Journal*, *23*, 275–284.

Caves, R. E., & Porter, M. E. (1977). From entry barriers to mobility barriers: Conjectural decisions and contrived deterrence to new competition. *Quarterly Journal of Economics*, *91*, 241–261.

Chadwick, C., Hunter, L. W., & Walston, S. L. (2004). Effects on organizational performance of human resource management practices used in lay-offs. *Strategic Management Journal*, *25*, 405–427.

Chan, D. (1998). Funtional relations among constructs in the same content domain at different levels of analysis: A typology of composition models. *Journal of Applied Psychology*, *83*, 234–246.

Chandler, A. D. (1962). *Strategy and Structure: Chapters in the History of the Industrial enterprise*. Cambridge, MA: MIT Press.

Cohen, W., & Levinthal, D. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, *35*, 128–152.

Dess, G., & Beard, D. (1984). Dimensions of organizational task environments. *Administrative Science Quarterly*, *29*, 52–73.

Dierickx, I., & Cool, K. (1989). Asset stock accumulation and sustainability of competitive advantage. *Management Science*, *35*(12), 1504–1513.

DiMaggio, P. J., & Powell, W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, *48*, 147–160.

Dyer, J. H., & Singh, H. (1998). The relational view: Cooperative strategy and sources of interorganizational competitive advantage. *Academy of Management Review*, *23*(4), 660–680.

Ettlie, J. E., & Sethuraman, K. (2002). Locus of supply and global manufacturing. *International Journal of Operations & Production Management*, *22*(3), 349.

Frost, T. S., Birkinshaw, J. M., & Ensign, P. C. (2002). Centers of excellence in multinational corporations. *Strategic Management Journal*, *23*, 999–1018.

Gersick, C. (1991). Revolutionary change theories: A multilevel exploration of the punctuated equilibrium paradigm. *Academy of Management Review*, *16*(1), 10–37.

Hambrick, D. C., & Mason, P. A. (1984). Upper echlons: The organization as a reflection of its top managers. *Academy of Management Review*, *9*, 193–206.

Hannan, M. T., & Freeman, J. (1984). The population ecology of organizations. *American Journal of Society*, *82*, 929–964.

Hawawini, G., Subramanian, V., & Verdin, P. (2003). Is performance driven by industry or firm-specific factors?: A new look at the evidence. *Strategic Management Journal*, *24*, 1–16.

House, R., Rousseau, D. M., & Thomas-Hunt, M. (1995). The meso-paradigm: A framework for integration of micro and macro organizational behavior. In: L. L. Cummings & B. M. Staw (Eds), *Research in organizational behavior*, Vol. 17 (pp. 71–114). Greenwich, CT: JAI Press.

Huff, A. S. (1982). Industry influences on strategy reformulation. *Strategic Management Journal*, *3*, 119–131.

Jensen, M. C., & Meckling, W. F. (1976). Theory of the firm: Managerial behavior, agency costs, and ownership structure. *Journal of Financial Economics*, *3*, 305–360.

Jensen, M. C., & Zajac, E. J. (2004). Corporate elites and corporation strategy: How demographic preferences and structural position shape the scope of the firm. *Strategic Management Journal*, *25*, 507–525.

Klein, K. J., Dansereau, F., & Hall, R. J. (1994). Levels issues in theory development, data collection and analysis. *Academy of Management Review*, *19*, 105–229.

Klein, K. J., Tosi, H., & Cannella, A. A. (1999). Multilevel theory building: Benefits, barriers, and new development. *Academy of Management Review*, *24*, 243–248.

Kozlowski, S. W. J., & Klein, K. J. (2000). A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. In: K. J. Klein & S. W. J. Kozlowski (Eds), *Multilevel theory, research and methods in organizations: Foundations, extensions and new directions* (pp. 3–90). San Francisco: Jossey-Bass Inc.

Krishnan, R. R., Joshi, S., & Krishnan, H. (2004). The influence of mergers on firms' product-mix strategies. *Strategic Management Journal*, *25*(6), 587–611.

Lee, J., Lee, K., & Rho, S. (2002). An evolutionary perspective on strategic group emergence: A genetic algorithm-based model. *Strategic Management Journal*, *23*, 727–747.

Levinthal, D., & March, J. G. (1993). The myopia of learning. *Strategic Management Journal*, *14*, 95–112.

Levitt, B., & March, J. G. (1988). Organizational learning. *Annual Review of Sociology*, *14*, 319–340.

Luo, Y. (2002). Product diversification in international joint ventures: Performance implications in an emerging market. *Strategic Management Journal*, *23*, 1–20.

Madhok, A. (2002). Reassessing the fundamentals and beyond: Ronald Coase, the transaction cost and resource-based theories of the firm and the institutional structure of production. *Strategic Management Journal*, *23*, 535–551.

Miles, R. E., & Snow, C. C. (1978). *Organizational strategy, structure, and process*. New York: McGraw-Hill.

Miller, K. D. (2002). Knowledge inventories and managerial myopia. *Strategic Management Journal*, *23*, 689–706.

Morgeson, F. P., & Hofmann, D. A. (1999). The structure and function of collective constructs: Implications for multilevel research and theory development. *Academy of Management Review*, *24*(2), 249–265.

Park, N. K., Mezias, J. M., & Song, J. (2004). A resource-based view of strategic alliances and firm value in the electronic marketplace. *Journal of Management*, *30*, 7–29.

Parkhe, A. (1993). Strategic alliance structuring: A game theoretic and transaction cost examination of interfirm cooperation. *Academy of Management Journal*, *36*(4), 794.

Peng, M. W. (2004). Outside directors and firm performance during institutional transitions. *Strategic Management Journal*, *25*, 453–471.

Penrose, E. (1959). *The theory of growth of the firm*. London: Basil Blackwell.

Peteraf, M. A. (1993). The cornerstones of competitive advantage: a resource-based view. *Strategic Management Journal*, *14*(3), 179–191.

Pfeffer, J., & Salancik, G. R. (1978). *The external control of organizations*. New York: Harper and Row.

Porter, M. (1980). *Competitive strategy*. New York: Free Press.

Porter, M. (1981). The contributions of industrial organization to strategic management. *Academy of Management Review*, *6*, 609–620.

Porter, M. (1990). *The competitive advantage of nations*. New York: Free Press.

Porter, M. (1998). Clusters and the new economics of competition. *Harvard Business Review*, *76*(6), 77–90.

Pouder, R., & St. John, C. (1996). Hot spots and blind spots: Geographical clusters of firms and innovation. *Academy of Management Review*, *21*(4), 1192–1225.

Prahalad, C. K., & Bettis, R. A. (1986). The dominant logic: A new linkage between diversity and performance. *Strategic Management Journal*, *7*, 485–501.

Priem, R. L., & Butler, J. E. (2001). Is the resource-based view a useful perspective for strategic management research? *Academy of Management Review*, *26*, 1–22.

Priem, R., Lyon, D., & Dess, G. (1999). Inherent Limitations of demographic proxies in top management team heterogeneity research. *Journal of Management*, *25*(6), 935–953.

Ramos-Rodriquez, A. R., & Ruiz-Navarro, J. (2004). Changes in the intellectual structure of strategic management research: A bibliometric study of the strategic management journal, 1980–2000. *Strategic Management Journal*, *25*(10), 981–1004.

Roberts, K. H., Hulin, C. L., & Rousseau, D. M. (1978). *Developing an interdisciplinary science of organizations*. San Francisco: Jossey-Bass.

Rodan, S., & Galvnic, C. (2004). More than network structure: How knowledge heterogeneity influences managerial performance and innovativeness. *Strategic Management Journal*, *25*, 541–562.

Rousseau, D. M. (1985). Issues of level in organizational research: Multilevel and cross-level perspectives. In: L. L. Cummings & B. M. Staw (Eds), *Research in organizational behavior*, (Vol. 7, pp. 1–37). Greenwick, CT: JAI Press.

Rouse, M. J., & Daellenbach, U. S. (1999). Rethinking research methods for the resource-based perspective: Isolating research of sustainable competitive advantage. *Strategic Management Journal*, *20*(5), 487–494.

Rumelt, R. P. (1974). *Strategy, structure, and economic performance*. Cambridge, MA: Harvard University Press.

Rumelt, R. P. (1984). Towards a strategic theory of the firm. In: R. B. Lamb (Ed.), *Competitive strategic Management* (pp. 556–570). Englewood Cliffs, NJ: Prentice-Hall.

Sakakibara, M. (2002). Formation of R&D consortia: Industry and company effects. *Strategic Management Journal*, *23*, 1033–1050.

Schendel, D., & Hofer, C. (1979). In: Dan Schendel & Charles Hofer (Eds), *Strategic management: A new view of business policy and planning*. Boston: Little Brown.

Spender, J. C. (1989). *Industry recipes: The nature and source of managerial judgements*. Oxford, England: Basil Blackwell.

Thompson, J. D. (1967). *Organizations in action*. New York: McGraw-Hill.

Wernerfelt, B. (1984). A resource-based view of the firm. *Strategic Management Journal*, *5*(2), 171–180.

Wiersema, M. F., & Bantel, K. A. (1992). Top management team demography and corporate strategic change. *Academy of Management Journal*, *35*, 91–121.

Wilk, E. O., & Fensterseifer, J. E. (2003). Use of resource-based view in industrial cluster strategic analysis. *International Journal of Operations & Production Management*, *23*, 995–1010.

Williamson, O. E. (1975). *Markets and hierarchies: Analysis and antitrust implications*. New York: Free Press.

Williamson, O. E. (1985). *The economic institutions of capitalism: Firms, markets, relational contracting*. New York: Free Press.

Williamson, O. E. (1999). Strategy research: Governance and competence perspectives. *Strategic Management Journal*, *20*, 1087–1109.

Wright, P., Ferris, S. P., Sarin, A., & Awasthi, V. (1996). The impact of corporate insider, blockholder, and institutional equity ownership on firm risk taking. *Academy of Management Journal*, *39*(2), 441–464.

Yammarino, F. J., & Dubinsky, A. J. (1994). Transformational leadership theory: Using levels of analysis to determine boundary conditions. *Personnel Psychology*, *47*(4), 787–802.

CAUSE MAPPING IN STRATEGIC MANAGEMENT RESEARCH: PROCESSES, ISSUES, AND OBSERVATIONS

Devi R. Gnyawali and Beverly B. Tyler

ABSTRACT

Our primary objective is to provide method-related broad guidelines to researchers on the entire spectrum of issues involved in cause mapping and to encourage researchers to use causal mapping techniques in strategy research. We challenge strategists to open the black box and investigate the mental models that depict the cause and effect beliefs of managers, "walk" readers through the causal mapping process by discussing the "nuts and bolts" of cause mapping, provide an illustration, and outline "key issues to consider." We conclude with a discussion of some promising research directions.

Cause mapping is considered the most popular compositional technique for analyzing strategic judgment processes (Huff, 1990b; Priem & Harrison, 1994). While it is not a new concept (e.g., Axelrod, 1976; Toulmin, 1948), a general lack of methodological guidelines and skilled researchers to apply

Research Methodology in Strategy and Management
Research Methodology in Strategy and Management, Volume 2, 225–257

ISSN: 1479-8387/doi:10.1016/S1479-8387(05)02010-2

the technique and methodological challenges have limited its use in organization theory and strategic management (Porac & Thomas, 1989; Stubbart, 1989). In the late 1980s and early 1990s, there was a general call for more research using this and other cognitive mapping techniques in organizational settings (e.g., Porac & Thomas, 1989; Stubbart, 1989; Weick & Bourgon, 1986). This call led to some edited books and special journal issues that illustrated the use of cognitive mapping techniques (e.g., Huff, 1990a; Eden, Ackermann, & Cropper, 1992) and efforts to blend qualitative and quantitative techniques (Langfield-Smith & Wirth, 1992; Laukkanen, 1992; Markoczy & Goldberg, 1995). Still, relatively few articles using cause mapping have been published in mainstream U.S. management journals to date, although a few edited books have followed this early push (e.g., Meindl, Stubbart, & Porac, 1996; Eden & Spender, 1998). Furthermore, cognitive research methodologies are often not incorporated into mainstream strategy doctoral programs, many publications tend to emphasize a particular methodology or software, and very few publications attempt to comprehensively discuss the choices involved when conducting cause mapping research (i.e., issues from predata collection to data analysis and interpretation). The purpose of this chapter is to provide an overview for researchers wanting to know more about cause mapping and to discuss various method-related tradeoffs required when conducting cause mapping. Our hope is to encourage empirical research using cause mapping methodologies.

In this chapter, our objectives are limited. We do not provide a comprehensive review of the cause mapping literature. We will not discuss the theoretical issues that might lead to the use of cause mapping methodologies nor will we do an extensive discussion of a particular cause mapping approach or software. What we hope to accomplish is more modest. Our primary objective is to provide method-related broad guidelines to researchers on the entire spectrum of issues involved in cause mapping and to encourage researchers to use cause mapping techniques in strategy research and practice. In order to accomplish this, we will first discuss the growing recognition in the strategy literature of the importance of understanding how and why organizations and managers in them make the choices they do. We argue that cognitive researchers and managers are no longer satisfied with studies of strategic choice based only on managerial demographics or simplified strategy-making processes, which depict managers as simple-minded information processors. We challenge strategists to open the black box and investigate the mental models that depict the cause and effect beliefs of managers that influence their strategic choices, using cause mapping techniques. We will then "walk" readers through the entire process of cause

mapping and raise several salient issues. Next, we will provide an illustration of our recent cause mapping research and discuss the "nuts and bolts" of cause mapping and the ways in which managers' gestalt views and shared cognition at the organization level could be captured. We follow this with a simple critique of some of the cause mapping techniques used and outline "key issues to consider" when conducting cause mapping research. We conclude with a discussion of some promising research directions.

COGNITIVE MAPPING AND STRATEGIC MANAGEMENT

Cognitive researchers seek to open the black box of strategic choice by evaluating managers' judgments rather than limiting observation to managerial characteristics or strategy-making processes (Priem & Harrison, 1994). They are not satisfied with accepting measures of managerial characteristics as substitutes for measures of a more cognitive nature (e.g., Markoczy, 1997). They argue that managers are not simple-minded information processors whose biases or implicit theories can be captured by simple demographic characteristics (Walsh, 1988). Priem and Harrison (1994) grouped several conceptually and technologically distinctive methods for exploring individual judgments into two categories: decompositional and compositional methods. They argued that a decompositional methods, such as metric conjoint analysis and policy capturing, present salient strategy variables to assess managers' preference judgment in response to each combination (judgment focus), while a compositional method gains insight into the process used to 'walk through' a decision-making situation and the variables considered in making a decision (process focus). They mentioned that compositional methods, such as cause mapping are particularly useful for theory building and argue that these process-focused methods attempt to identify mediating processes "between the perception of strategic variables and the development of a strategic judgment" (Priem & Harrison, 1994, p. 318).

Cognitive researchers have established a generally accepted understanding of the basic structure of an individual's cognitive understanding or knowledge and how this can be presented graphically (e.g., Axelrod, 1976; Bougon, Weick, & Binkhorst, 1977; Haray, Norman, & Cartwright, 1965). Research in organizational cognition suggests that managers' schema or mental models reflect their cognitive understanding of a particular domain

(Jelinek & Litterer, 1994; Stubbart & Ramaprasad, 1990). A schema is defined as a knowledge framework that selects and actively modifies experience in order to arrive at a coherent, unified, expectation-confirmation and knowledge-consistent representation of experience (e.g., Alba & Hasher, 1983). A managerial schema is a structured, domain-specific pack of knowledge that allows the manager to interpret, encode, store, and retrieve information appropriately to make various decisions (e.g., Lord & Foti, 1986).

A commonly accepted representation of knowledge is a semantic network model or a cognitive map (Huff, 1990b; Weick & Bourgon, 1986). It is argued theoretically that information is stored in the long-term memory in terms of an associated network that consists of concepts and the linkages between them. In these maps, concepts are referred to as "nodes" and the linkages between them are called "arcs." The concepts with the most linkages are argued to be more important and better integrated into the individual's mental understanding. These concepts are considered core or superordinate concepts and are more central in their cognitive network than concepts with fewer links (Rosch, 1973). Peripheral or subordinate concepts are included in the individual's mental understanding, yet are less integrated with other concepts and are less central. The linkages between concepts can vary in strength, i.e., a concept may be strongly connected with some concepts and moderately or weakly connected with others. Information in a knowledge network is accessed as an individual mentally (consciously or unconsciously) follows the linked path from one concept to another. The stronger the association between concepts, the more the thinking of one concept will trigger the thought of the other (Anderson, 1983). The strength of association between an activated node and all linked nodes determines the "spreading activation" and the specific information retrieved from memory (Collins & Loftus, 1975; Keller, 1993).

In the case of cause maps, the linkages between concepts in a cognitive map represent a causal link. If two concepts are positively linked, it is argued that an increase in the first concept influences or causes an increase in the second concept. If two concepts are negatively linked, an increase in the first concept is believed to lead, influence, or cause a decrease in the second concept. If no link is depicted, the person whose cognitive understandings have been mapped does not see an important causal association between the two concepts.

Thus, it has been argued that the ability of executives to receive, store, retrieve, and use appropriate information from their knowledge network will depend on (a) the concepts currently included in the knowledge framework (the content of their knowledge) and (b) the linkages among these

concepts (the structure of their knowledge: strength and causality). This in turn will depend on the content and structure of the knowledge previously stored and the new information being processed (Hitt & Tyler, 1991; Tyler & Steensma, 1998).

Therefore, a cause map is a specific form of cognitive map that incorporates concepts linked together by causal relationships. Although some people have described cognitive maps and cause maps in a similar manner (see discussion by Weick & Bourgon, 1986), we focus specifically on cause maps. Huff suggested that cause mapping, which is one of the five generic families of cognitive mapping methods she discussed, is "the most popular mapping method in organization theory and strategic management" (Huff, 1990a, p. 16). She noted that causal maps are based on some unique assumptions about cognition: (1) causal associations are the major ways in which understanding about the world is organized, (2) causality is the primary form of post hoc explanation of events, and (3) choice among alternative actions involves causal evaluation.

Weick (1990) argued that while traditionally maps have emphasized spatial relatedness, the interesting thing about strategic maps in management is that they also seem to capture time as they portray causality, predictive logic, or consequences. They also emphasize classification and assignment of things to classes, which allows managers and researchers to see patterns interspersed among the differences. According to him, accuracy is not always crucial in managerial maps; although accuracy is agreeable, it is not necessary for the organization to generate action given that "strategy implementation is often judged successful when the organization is moving roughly in the same direction" (Weick, 1990, p. 6). Weick writes the following:

> The important feature of a cause map is that it leads people to anticipate some order 'out there'. It matters less what particular order is portrayed than that an order of some kind is portrayed.... The map animates managers, and the fact of animation, not the map itself, is what imposes order on the situation. (Weick, 1985, p. 127)

In conclusion, causal mapping depicts the causal relationships managers use to order their thought processes as they make strategic decisions. They portray causality and uncover mediating processes, providing predictive logic for consequences. In doing so, mental maps capture time and emphasize classification and categorization, which allows managers and researchers to see patterns. These perceived patterns influence managers' actions in important ways. Thus, individual, group, and organizational causal attributions are imperative to the understanding of strategic choice within

complex organizations. Responding to what we see as a need for cognitive research using cause mapping techniques in strategic management, we provide an overview of the cause mapping process. After that, we will describe a recent study to illustrate mapping issues, discuss "key issues to consider" when conducting cause mapping research, and propose some promising directions for future research.

THE CAUSE MAPPING PROCESS

In this section, we first provide a simple overview of the steps involved in planning and executing cause mapping (Fig. 1). Then we break down the various steps into three major stages: predata collection, data collection and preparation of cause maps, and postdata collection. While we seek to be thorough in providing a generic discussion of the stages, it is important to note that the specifics of various steps vary slightly depending on the research context, purpose of research, researcher preferences, and other factors. Our general description of these steps will be clearer when we discuss our illustrative study.

As noted in Fig. 1, cause mapping research begins with the selection of a specific domain of cognition, a clear research purpose, and specific research questions (step 1). Next, researchers must determine what type of data will be collected and used (published documents or solicitation of data from participants), and the method of participant selection, if data are to be solicited (step 2). Once researchers understand the kind of data to be used, they need to prepare guidelines for data coding, construct and administer instruments if solicitation is required, and code the data (step 3). Using the coded data, researchers should be able to identify important concepts and group them (step 4) and identify cause–effect relationships in the form of cause maps (step 5). Maps can then be validated through discussions with executives or participants (step 6) and then analyzed (step 7). Finally, results can be reported, inferences drawn, and actions proposed and implemented (step 8). These eight steps can be organized within three stages of the cause mapping process.

Stage 1. Predata Collection Stage

Key issues in the predata collection stage involve specification of the domain of the study, the purpose of mapping, and the research questions. Similarly,

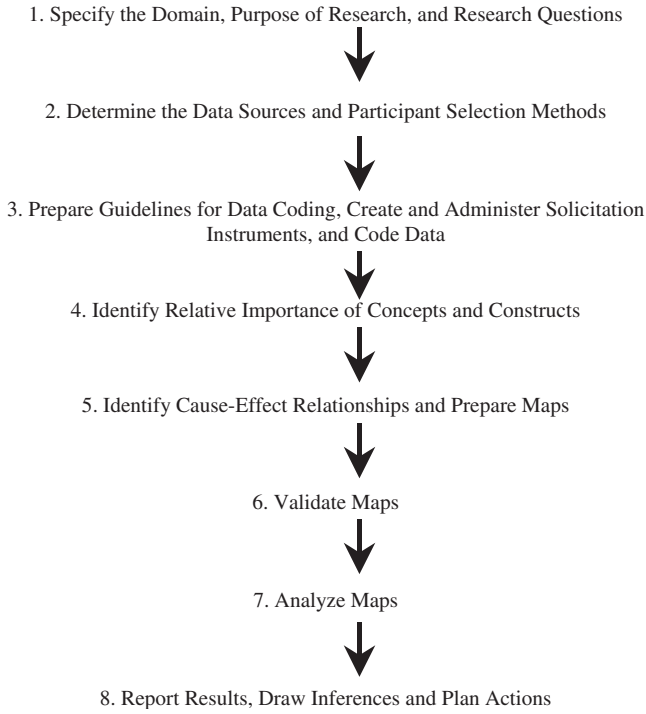

Fig. 1. An Overview of the Cause Mapping Processes.

it is important to be clear on the research context, sources of data, unit of observation, and the selection of the participants (if using primary data sources). Specifics of these steps vary depending on the domain of investigation.

Specify the Domain, Purpose, and Research Questions

The selection of the cognitive domain to be investigated is very important because the human mind has innumerable domains, and a researcher must provide clear limits on the domain under consideration. For example, are you trying to capture managers' cognitive understanding of their company's market orientation or critical issues involved in new product development? Or, is the goal to capture a manager's view of a specific problem situation? Although this is a generic step in any research project, it is more critical in cause mapping because cognitive maps are very domain specific. Lack

of specification of the domain will cause problems in data collection and analysis and could result into misleading conclusions. The clearer the researcher is on the domain one is trying to map or capture, the more precise the researcher can be in subsequent steps. Once you have decided on the domain of knowledge you wish to tap, you must determine the purpose. Are you developing theory or testing theory? Are you trying to diagnose problems and develop interventions or improve decision making? Finally, what are your specific research questions? What are you looking for or what do you expect to find? Are you comparing managers' cause maps within a business or across businesses? Are you evaluating the current situation or looking at change over time?

Research Context, Data Sources, Unit of Observation, and Participant Selection

The specific context selected for a study is driven by the domain, purpose, and research questions. However, even within these constraints, researchers make choices; e.g., between industries, companies, products, etc. When making the final choice there should be a very clearly defined logic regarding the context selected, tied as closely to prior research as possible (e.g., similar context to prior work or purposely different context for comparison). In more qualitative research, such as cause mapping, theoretical rigor is often scrutinized even more carefully than in empirical research, where the methods and data analysis may attract the attention of a reviewer. Decisions on the data source will depend to a large extent on the domain of study and research questions. If the researcher is interested in understanding cause maps of real managers and wants to compare them across functions and levels, it is better to collect data through interviews of these managers (primary sources). On the other hand, if the researcher is interested in understanding how a cause map of a company's senior management has changed over time, one could perform content analysis by coding the company's Annual Reports (or similar official company documents) for several years. Researchers planning to capture individual cognitive maps of various managers and aggregating them at the departmental, functional, or company levels need to carefully consider the issues of data aggregation (discussed later in the chapter). Also, if the goal is to generalize at the organization level through the cause maps of selected managers, the researcher needs to make sure that sampling of the managers is appropriately done to ensure that the sample does represent the various levels and functions of the organization.

Stage 2. Data Collection and Map Preparation Stage

The processes of data collection and representation of cause maps are certainly the most critical and time consuming ones in the cause mapping processes. We have included four of the eight steps noted in Fig. 1 as a part of this stage.

Coding Guidelines, Data Collection, and Data Coding

The first step is to develop coding guidelines, then collect the data, and code it. It is critical that researchers develop a detailed coding scheme that will capture the concepts of interest and categorize them in a way consistent with the study's domain, purpose, and research questions. The primary goal here is to have a useful coding scheme that can be used by two or more coders to reliably code the data (Miles & Huberman, 1984). While it is important to have the coding scheme reviewed and validated before starting to code the data, the coding scheme can also be modified (to increase its relevance and validity) during the data collection process as researchers understanding becomes more complete. The actual number of concepts researchers include in the guidelines to be coded varies greatly depending on the domain of interest, the number and type of documents coded, mapping issues involved (i.e., complexity of causal relationships), etc.

Various ways of data collection are found in the literature. The most common include post hoc coding of documents, interviews, written solicitation instruments, and direct brainstorming. The document-coding approach requires only the development of a detailed coding scheme, which will be used in the actual coding process. Barr and Huff (1997) and Narayanan and Fahey (1990) coded published documents to generate concepts which they then mapped. The other approaches require the development of instruments and detailed procedures for primary data collection, as well as coding guidelines. Calori, Johnson, and Sarnin (1994) used open-ended interviews with guiding questions to collect executives' views on the domain of their interest (changes in the environment and effects of such changes on the firm). Tyler and Gnyawali (2002) used interviews to generate concepts relevant to their domain of interest (managers' views of their firm's market orientation), and validated the concepts by examining relevant literature before administering written instruments to solicit managers' causal attributions. Finally, using softwares like Decision Explorer, researchers could collect data in group brainstorming sessions or individually with the participants (Eden et al., 1992).

Some researchers generate cause maps directly by using the first-order concepts (e.g., Cossette, 2002; Eden et al., 1992), whereas others create higher order categories or constructs by combining the first-order concepts (Laukkanen, 1994; Tyler & Gnyawali, 2002). If the researcher's interest is to capture and represent a large number of concepts in a map (generate a very complex maps consisting of many concepts), then creating constructs may not be part of the mapping process. However, researchers should recognize that this level of complexity may limit the reliability of coding across coders. Moreover, if researchers are using software, they could manage a large number of concepts. If not, it may be better to create higher-order constructs and prepare maps based on such constructs. That way, researchers have fewer constructs in the map and the maps can be more easily compared across various levels and units of the organization over time.

The researcher can adopt different approaches to create constructs for mapping. Laukkanen (1994) standardized the concepts into constructs through his own interpretation of the idiosyncratic concepts of various individuals. Tyler and Gnyawali (2002) performed qualitative content analysis of the raw concepts and merged the similar concepts into theme-oriented constructs. If mapping is done in a group environment, the researcher could first conduct a brainstorming session to generate the concepts, and lead the group discussion to facilitate the creation of higher-order constructs by the participants themselves. Research to date suggests that it is advisable to keep the number of constructs to a manageable number. For example, Markoczy and Goldberg (1995) suggest limiting the number of concepts to 10, while Tyler and Gnyawali (2002) presented participants with 20 concepts, but gave them discretion in dropping and adding concepts. While we do not propose a minimum or maximum limit here, we do suggest the researcher to carefully evaluate the trade-off between completeness in concepts and ability to generalize and compare the data across subjects (Weick, 1979, p. 35–41).

Identify Relative Importance of Concepts and Constructs

As noted earlier, researchers differ in term of using the first-order concepts or higher-order constructs. However, the way in which relative importance is assessed is similar. To assess the relative importance, researchers could ask the participants to either rank-order the concepts or to rate their importance on some kind of rating scale (e.g., Gnyawali, 1997). One could also assess managers' explicit statement about the importance of various constructs during conversations. Even if constructs are created by merging concepts, relative importance of concepts could be assessed if the researcher

is interested in finding out specific details at the concept level as well. For example, although Calori et al. (1994) created constructs from the first-order concepts they generated, they still assessed the relative importance of the concepts through four criteria: (1) managers' explicit mention of the importance of the concepts, (2) spontaneity in mentioning the concept, (3) priority accorded to the concept during the interview, and (4) relative length of time spent in discussing the concept.

Identify Cause–Effect Relationships and Prepare Maps

Researchers have adopted various approaches for this purpose. Identification of cause and effect relationships in published documents (secondary data) must be done by the coders following the coding guidelines. While terms suggesting a cause and effect relationship should be included in the guidelines, judgments may be required when causation is implied. Rules of how this should be handled need to be specified. Hodgkinson, Maule, and Brown (2004) have compared the two most common approaches used in primary data analysis (free-hand and pair-wise evaluation). In the free-hand approach, participants are presented with a list of constructs or concepts and asked to identify any cause and effect relationships among any or all the concepts by drawing links connecting the cause (denoted by the origin of an arrow) and effect (denoted by the destination of an arrow) relationship among the concepts. In the pair-wise approach, participants are normally provided a paired list of concepts and they are asked to consider if there is a cause and effect relationship between every pair of the concepts. This process is continued until the participant has considered all possible pairs and noted all the cause and effect relationships. An adjacency matrix (with $n \times n-1$ cells to indicate cause–effect relationships) is often provided to ease the identification of cause and effect relationships.

A key part of the cause–effect identification is a specification of the nature (positive or negative) of relationship between the concepts. Two concepts are positively related if they move in the same way, i.e., an increase (or decrease) of the first concept leads to increase (or decrease) in the second concept. The identification of positive or negative relationships is an important step, especially if interventions are to be designed based on managers' cognitive maps. Some researchers also identify the strengths of relationships (usually in a scale of 1–3).

Once the cause and effect relationships are identified, the next step is to prepare a visual representation of the relationships. The primary task here is to visually depict the concepts or constructs and cause–effect relationships identified in the prior steps in the form of a cause map (see Fig. 2). Usually,

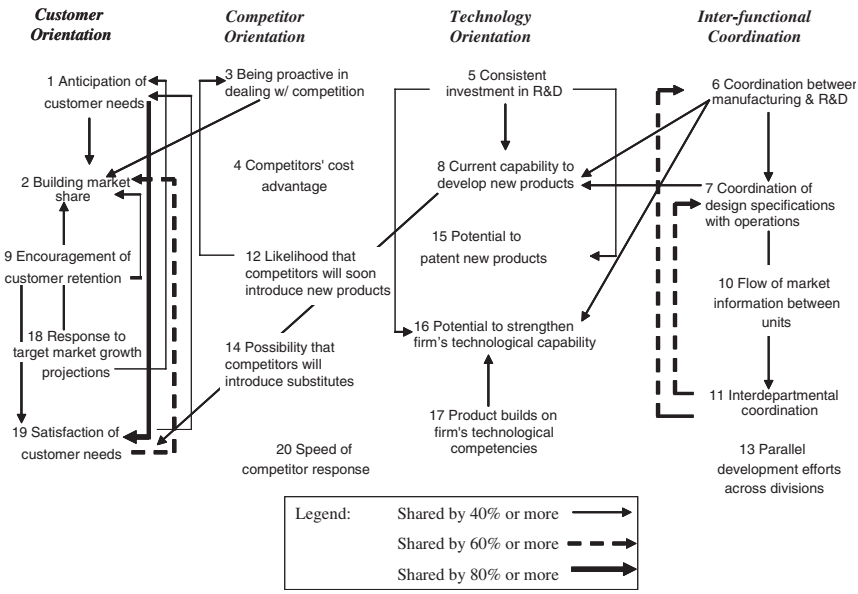

Fig. 2. An Illustrative Cause Map.

maps have some kind of organizing scheme to make it easier to grasp and interpret the map. Two commonly used approaches are the themes reflected by the concepts (e.g., Calori et al., 1994; Tyler & Gnyawali, 2002) and means-givens-ends as reflected by the cause and effect relationships (Fuglseth & Gronhaug, 2002). In the former approach, all concepts that relate to a particular theme are grouped together (for example, the theme of customer orientation in Fig. 2 is on the extreme left of the map). In the latter approach, means (concepts that have the most number of out arrows) are grouped on the extreme left, followed by givens (concepts having same number of in and out arrows) in the middle, and ends (concepts having the most number of in arrows) are grouped on the extreme right of the map.

Validate Maps

Before the maps are aggregated, analyzed, and final conclusions are drawn, it is important that the researcher makes sure that the maps prepared are valid. If mapping is done using primary data sources, the researcher could go back to the participants, share the initial maps, and get their feedback. If the maps are based on secondary data, the researcher could talk to some

experts, including company managers, to make sure that the constructs and the relationships are valid. Several aspects of validation are important. For example, does the pool of concepts adequately represent the domain, i.e., are the concepts or constructs complete? If not, should any new ones be added? Are the constructs derived through the aggregation of concepts meaningful? Do the concepts inside the constructs make sense, i.e., does a particular concept in a construct belong there and nowhere else? Are the cause and effect relationships drawn appropriate? Is something misrepresented or mistaken? If maps are interpreted or inferences are drawn, do the inferences make sense? Are they all focused on the domain of interest?

Stage 3. Postdata Collection Stage

Analyze Maps

Several things are involved in analyzing the maps. Examples include aggregation, comparison of maps (if necessary), assessment of centrality of concepts and complexity of maps, and identification of the themes of the maps. If the researcher is interested in creating one or more shared maps of various individuals and units of the organization, then it is important that the maps be aggregated. Weick and Bourgon (1986) suggest three ways to prepare shared maps: composite, assemblage, and average. Preparation of a composite map requires all participants to come together, discuss, and develop one common map. The assemblage method involves preparation of the map of a dominant department or unit. This method is not relevant if researchers are interested in examining the level of shared meaning among managers, because only one dominant map is created. The average method is useful for researchers interested in capturing the extent of shared cognition. In this method, individual adjacency matrices of all the managers are averaged to compute shared maps (Gnyawali, 1997; Tyler & Gnyawali, 2002).

Another aspect of map analysis may involve comparison of maps. Maps can be compared in several ways: (a) comparison of contents (at both concept and construct level) to examine if the same or similar concepts and constructs exist in various maps; (b) comparison of links to examine the existence of links between various pairs of concepts or constructs; and (c) comparison of strength and directionality of links to examine if the links are weak or strong, and positive or negative. Map comparison may be more qualitative in nature as visual differences in the maps are evaluated in a discussion format (Barr, Simpert, & Huff, 1992). However, one could also use computer software, such as CMAP (Laukkanen, 1998) to compare

various maps. Comparison would then be made based on standardized concepts (researcher-identified constructs or categories), not on idiosyncratic concepts generated directly through the interviews or coded documents. An even more quantitative way to compare various maps is the distance ratio and related analyses proposed by Markoczy and Goldberg (1995).

The kind of comparison to be done depends on the research question and related factors. Commonly used approaches are comparison across levels (individual, group, organization) and functions of the organization (Tyler & Gnyawali, 2002) and comparison across individuals at the same time period (Stubbart & Ramaprasad, 1988). Other approaches include comparison of the same individual, group, organization, or function over multiple periods (Narayanan & Fahey, 1990; Huff & Schwenk, 1990), comparison of maps before and after an intervention, such as organizational learning processes (Gnyawali, 1997), and comparison of participant map with the expert map (Gnyawali, 1997).

Most researchers use a combination of qualitative and quantitative techniques to analyze the maps. Key aspect of analysis include the following: (a) identification of means-givens-ends or causes and consequences; (b) identification of themes or dominant aspects of the maps; (c) assessing the importance of constructs (described above); (d) cluster analysis and multidimensional scaling to examine how concepts group together; and (e) computing density of the maps and centrality of concepts. Network analysis software such as UCINET (Borgatti et al., 1999) could be used to analyze centrality and density.

Report Results, Draw Inferences, and Plan Actions

Laukkanen (1998) suggests that researchers need to pay careful attention to how they present results from causal mapping. This is very important because reporting standards for cognitive mapping are not well developed (unlike in research using quantitative techniques). Researchers should be careful not to overburden the reader with many complex maps or be carried away with fancy presentation of maps without clear insights articulated based on the maps. Therefore, a key concern of the researcher should be the key message and insights generated from the study and conveyed by the maps. In some respects, drawing inferences and reporting them requires the skills of an artist, as the researcher follows the causal implications reflected in the maps and draws on information they have obtained in early interviews, pilot tests, and feedback sessions with participants. It is helpful as well to ask colleagues' feedback on various graphic appearances of the maps before making the final decision regarding how to portray the maps and

discuss insights from them (Laukkanen, 1998). If the research questions are related to learning or change, interventions may be proposed and initiated. In any event, some form of action plan should be proposed based on the results of the study and the specific research questions under investigation. Now let us illustrate these cause mapping steps.

AN ILLUSTRATION OF CAUSE MAPPING

This illustration is based on a recent study in which we examined managers' cognitive structures of their business's market orientation (MO) in the context of new product development (Tyler & Gnyawali, 2002). Here, we focus on the methodology used in the study and will only briefly discuss the results.

Stage 1. Predata Collection

The domain of cognitive understanding we wanted to study was how managers in a business conceptualize what it means to be market driven or MO in the context of new product success. Our purpose was to illustrate cause maps of managers from a company known for being MO and to contribute to theory development. We wanted to examine cause maps of the business unit as a whole and of various functions and levels of the business unit. Two of our research questions are relevant to this discussion. First, we wanted to know how a business's MO regarding new product success was reflected in managers' cause maps. Second, we were interested in learning how managers' cause maps of MO differed across levels and functions.

We carefully selected the industry and company for our study. Based on prior MO literature, we selected an industry that had moderate levels of market growth and technological turbulence in an effort to select a context where we would expect to find customer orientation emphasized over other dimensions of MO. The criteria of moderate industry growth and technological turbulence led us to select the frozen food (FF) industry. We selected a business from this industry known for being MO, because we wanted to investigate a business that exemplified the characteristics researchers have argued are important to being MO. The business selected was a major U.S. subsidiary of a multinational food company, which, according to its 1999 annual report, is the "undisputed leader in its sector" and holds first or second place in almost all of its product categories. The corporation and its U.S. FF division have a long history of focusing on the needs of customers

and are known for their ability to produce successful new food products. In the business, a real growth spurt began in 1998, when one of the most significant new product lines in its' history was introduced. This 1998 introduction of a new product line served as the context in this study. We conducted the study within a single business and focused on one major new product introduction, because prior research has shown that the effects of MO on performance are contingent on industry characteristics, organizational characteristics, product characteristics, and the measure of performance studied (e.g., Henard & Szymanski, 2001). Based on the literature, we reasoned that in this context we should find managers, across the business, share common knowledge structures regarding customers, competitors, technology, and inter-functional coordination (Tyler & Gnyawali, 2002). Because we chose to study managers at multiple levels and functions, secondary data were unavailable and primary data were solicited.

Adhering to the logic of theoretical sampling (Glaser & Strauss, 1967; Yin, 1994), the managers for this study were selected with the aim of collecting diverse range of perspectives on the phenomenon under investigation (Thomas, Sussman, & Henderson, 2001). Accordingly, a stratified sample of 40 managers, representing four levels (top management team, next two levels, and a forth level only in product development) and all major functional groups in the business (new venture/product development, marketing, operations, supply chain management, and finance) were asked by mail to participate in the study. We were interested in examining the extent to which cause maps are shared among managers within the business unit and within each level and function. We also wanted to examine unique differences in cause maps across levels and functions. Since our data were collected from individual managers, we had to aggregate the various cognitive maps in order to examine shared understanding.

Stage 2. Data Collection and Map Preparation Stage

A number of steps were required before the primary data could be collected. We created preliminary data-collection instruments, conducted initial interviews, coded the data collected during these interview to identify concepts, and grouped the concepts into constructs as reported in the MO literature, All of this was required in order to create the solicitation instruments and coding guidelines for the primary data to be collected. First, in an effort to better understand what being "market driven" meant in this business and identify preliminary concepts, we conducted initial

semi-structured interviews with managers at four levels ($N = 10$: the president, four executives that reported to him, and five in the next two levels of management). The managers were first asked to respond to open-ended questions related to being "market driven." Next, they were asked to provide up to ten words or phrases that described what being "market driven" meant to them. Finally, they were asked to provide up to ten words or phrases to describe each of the initially selected words or phrases. This is a fairly common solicitation technique used to capture cognitively related concepts.

The concepts generated during these interviews were content coded and grouped under the four primary constructs that have been suggested as major dimensions of MO (Gatignon & Xuereb, 1997): customer orientation (CU), competitor orientation (CP), technology orientation (TO) and interfunctional coordination (IC). From this list of concepts, we selected 20 phrases representing the four dimensions of MO (five for each dimension), to be used in the cause mapping exercise distributed to the final sample of participants. These 20 concepts or phrases were placed on 20 cards. These phrases and related cognitive mapping instruments were pre-tested with 15 graduate students experienced in new product development. Minor revisions were made to these instruments before conducting the study.

Every participant was mailed these 20 cards, two extra blank cards, instructions for sorting them, and an adjacency matrix with instructions for completing it. The step-by-step instructions asked the managers to identify the set of important concepts (they could select any or all the 20 concepts and could add any new important concepts by writing them on the blank cards), rank-order the relative importance of these concepts (i.e., rank from most to least important), and note cause-and-effect relationships between each concept with up to three other concepts in the adjacency matrix (i.e., positive or negative). Very few managers added any new concepts, suggesting that the 20 concepts proposed represented the most salient concepts in this cognitive domain. A total of 30 managers completed the cause mapping tasks as instructed. Data from these managers were used in subsequent analyses as described below.

Thus, the base concepts for the cause mapping data were collected through primary sources (interviews with managers), aggregated into four higher-order constructs according to the academic literature, and verified by the managers themselves. Cognitive maps were prepared based on the relationships managers identified among the 20 concepts (or fewer concepts depending on the managers' own view of MO). This approach provided standardization of the concepts used in cause mapping (making it easier to

compare maps across the managers and aggregate them) and yet provided opportunity for managers to customize the set of important concepts depending on their views.

In order to identify links, prepare the cause maps, validate raw maps, and aggregate the individual maps into the shared maps of the overall business and of various functions and levels, we prepared two data sets from the mapping data: one based on the rank-ordered data and the other on the cause-and-effect relationships in the adjacency matrices. The rank-ordered data were aggregated to create ranking results for each of the four dimensions of MO for various levels (Table 1a) and functions (Table 1b) of the business. The numbers in these tables were calculated by averaging the mean rankings of the five concepts for each dimension (please note that lower numbers indicate higher importance because the numbers are based on rank-ordered data). Similar to the ranking results, the adjacency matrices

Table 1. Average Ranking of MO Dimensions.

	N	Customer	Competitor	Technology	Inter-functional Coordination
(a) Levels					
Level 1	4	7.45	13.55	13.00	11.70
Report to Level 2	10	6.16	13.82	14.26	12.84
2 levels from Level 3	11	6.47	15.71	12.45	11.90
3 levels from Level 4 (product devt. specialists)	5	8.72	16.92	11.56	10.56
Total/overall	30	7.20	15.00	12.82	11.75
(b) Functions					
New Venture[a]	6	5.63	14.30	13.57	13.13
Marketing	5	5.48	14.48	14.16	10.12
Operations	6	8.63	15.63	11.90	10.77
Supply chain	4	5.60	13.30	13.05	13.10
Finance	4	6.80	15.00	13.80	11.55

[a]This group had 11 managers representing four levels, Since no other group had managers in level 4, These data are from six managers of levels 1–3.

were aggregated to create an overall adjacency matrix of all the 30 executives, and that matrix was used to prepare the overall cognitive map of the business (Fig. 2). In the cognitive map, concepts are grouped according to the four dimensions under investigation. The five concepts on the left end of the map relate to customers, the next five to competitors, and so on, as indicated across the top of each map. The arrows indicate cause (origin) and effect (destination) relationships among the concepts denoted by the executives. Each connecting arrow is counted twice, once for the cause and once for the effect. The level of agreement among executives on cause-and-effect relationships is reflected in the thickness of the arrows. The links shown in Fig. 2 are shared by at least 40% ($N = 12$) of the executives, and some links are shared by 60% (18 managers) or 80% (24 managers) of the executives as noted at the bottom of the cognitive map. Thus, the cause and effect relationships denoted are commonly shared across the business. Similar maps were constructed by aggregating individual data by level and function.

After initial aggregated maps were computed and illustrated, the maps were sent to the company sponsor and feedback was sought regarding the study's findings. An aggregated map of all participants was sent to each participant, along with personal feedback. Feedback suggested that the maps had face validity and no one voiced concerns about our interpretations of the results.

Stage 3. Postdata Collection

Our analysis focused on discerning the importance placed by managers on various concepts and dimensions of MO and comparisons of maps across levels and functions. Both the rank-ordered data and the cause maps reveal that the managers strongly emphasized the customer dimension of MO. The overall ranking results (the last row of Table 1a) clearly suggest that the managers viewed CU as being much more important than IC and TO, and that CP was viewed as the least important orientation. Moreover, an evaluation of overall ranking results of each of the 20 concepts used for mapping (not reported in the table) revealed that the top 3 ranked concepts as well as 4 out of the top 6 ranked concepts were all CU items. On the other hand, the 2 least important concepts and 4 out of 6 least important concepts were CP items.

The number of links in the overall cause map suggests yet another measure of the importance (from highest to lowest) of the various dimensions:

CU (18 links), TO (12 links), IC (11 links), and CP (3 links) (Fig. 2). Moreover, the links associated with CU concepts are more commonly shared than others, thus implying the importance of CU relative to CP, TO, and IC. The cause-and-effect links in this map suggest that the managers believe that the key goal (as indicated by the arrows heads) of this business is to "build market share." Other goals include "satisfaction of customer needs" and "anticipation of customer needs," both of which help to increase market share. The fact that all these key goals are customer related also shows the importance placed by the managers on CU.

To understand some of the similarities and differences in the knowledge structures of the managers across the levels and functions, we examined ranking results and cognitive maps of each level and function sampled. We now briefly describe results based on this analysis. In general, the rankings and cognitive maps were relatively consistent with the overall rankings and the map discussed above. For example, every level ranked CU as the most important, almost every level ranked CP as the least important, and every level ranked IC as more important than TO or CP (Table 1a). However, the second level managers ranked CP as more important than TO. The functional ranking results (Table 1b) also suggest that the managers viewed CU as being by far the most important, CP as the least important, and IC as more important than TO.

In terms of cause maps, the level 1 map revealed a sophisticated understanding of how CU, CP, TO, and IC concepts are related. The lower level maps did not reflect this comprehensive or integrated understanding. The complex and integrated nature of the level 1 map is consistent with the literature that suggests that senior managers have a more balanced and integrated cognitive understanding than do managers of lower levels (Lyles & Schwenk, 1992). The level 1 map also had many links that connect concepts across the four dimensions, suggesting that these senior managers believe that the MO dimensions influence each other as businesses engage in successful new product development activities. Most of the causal linkages recognized by the managers at the lower levels, however, are between concepts within a dimension rather than across the dimensions.

The five functional maps were very similar, but revealed some important differences. Managers of all functions had a more highly shared understanding regarding the customer dimension than other dimensions, and a very little shared understanding regarding the competitor dimension. All functional groups appeared to have a common understanding of how concepts in one dimension influenced or were influenced by those in other

dimensions (e.g., how concepts of IC dimension influenced those of TO dimension). However, the various functional groups tend to have very different understandings of the cause-and-effect relationships between the concepts. For example, while all five maps suggested that the managers shared a common understanding of the relationship between anticipation of customer needs and satisfaction of customer needs, the operations and supply chain managers believed that satisfaction leads to anticipation and the other functional managers believed anticipation leads to satisfaction, thus suggesting differences in perspectives regarding current versus future customer needs. The supply chain map was the most complex among all the functional maps.

A comparison of the functional maps to the level maps revealed two primary differences: (1) managers within a function seem to have more shared understanding than those within a level. On an average, 61.2 links existed in each functional map whereas the level maps had only 48 links; and (2) the links in each functional map included more cross-dimension links (e.g., TO concepts impacting CU concepts) than those in the level maps. On an average, each functional map had 10 cross-dimension links, whereas each level map had 7.5 such links.

Overall, our examination of cognitive maps of the managers of this business suggests the following regarding the business's MO. First, the dominant shared understandings of the business are that it is extremely important to (1) anticipate customer needs, (2) satisfy customer needs, and (3) build market share in order to be successful in new product development. This is illustrated by the higher-importance ranking of these specific concepts relative to other concepts and the existence of a larger number of shared links associated with these concepts across all maps. Second, if we consider the dimensions of MO rather than the individual concepts, their MO regarding new product success is customer focused. This contention is supported by the very-high-importance ranking of the CU dimension relative to the other dimensions and the existence of a larger number of shared links associated with the CU dimension than other dimensions across all maps. Furthermore, the managers believe that in order to serve customer needs and build market share, it is more important to have good coordination across various functions and to develop technological capability than worry about competitor actions. Finally, although the various levels and functions seem to share a common MO, there are still some differences. This is illustrated by the greater number of shared links within the functional maps relative to the level maps, and the differences in cause-and-effect links across the functions.

KEY ISSUES TO CONSIDER IN CAUSE MAPPING

Based on our above overview of the cause mapping processes and illustration of how to do cause mapping, we next outline some key issues to consider as researchers plan and implement cause mapping research. The primary issues we address are about data collection, data aggregation, tradeoffs involved in drawing inferences, and use of computer software. Table 2 provides a summary of key issues we discuss below.

Issues Related to Data Collection

As noted earlier, various options are available to collect data for cause mapping. Weick and Bourgon (1986, pp. 113–118) outline three main ways

Table 2. A Summary of Key Issues to Consider when Using Cause Mapping.

Areas for Consideration	Key Issues and Questions
Data collection	• What are the implications of using primary (interview based) versus secondary (published data based) sources of data? • Which data collection method suits the purpose of the research and research questions?
Data aggregation	• Does the research context call for the examination of shared cognition? • What are the various options to capture shared cognition? • What is an acceptable level of agreement?
Data presentation and drawing inferences	• What is an acceptable degree of complexity of the maps? • How to balance fanciness and practicality of the maps? • How to balance saliency and comparability of the maps? • How to increase reliability of the maps?
Use of cause mapping software	• What are the implications of using or not using software for cognitive mapping? • If a software is to be used, which software to use and why?

of gathering data for cognitive maps. They are systematic coding of documents representing the writings or statements of an individual, coding of verbatim transcripts of private meetings in which the individual participates, and eliciting causality beliefs through questionnaires and interviews. Huff and Fletcher (1990, pp. 406–407) also identify the option of interactively generating the data to be mapped and post hoc analysis of data generated for some other purposes. Huff (1990b) suggests that post hoc causal explanations have some advantages over primarily interview-based data collection. Post hoc data avoid the recall biases of interviews, are more detailed as the information about managerial thinking is rigorously collected, and provide insights into changing models overtime because post hoc coding could be done to cover an extended time period. Access to data is less of a problem with the post hoc coding and this method could be less time consuming because it is not necessary to spend a lot of time in interviewing and transcribing the interview, and future researchers could use the original data sources (enabling replication, if desired).

However, there are some obvious advantages of using the primary data approach. First, interviews and self-administered questionnaires allow the collection of data relevant to the researchers' purpose. Moreover, with interviews, the researcher is more likely to generate first-hand rich data through conversations with the participants, is more likely to capture underlying cognitions of the participants, and the data is more likely to be reliable. Most importantly, the direct interview method provides a two-way communication process, and a person's cognition is better captured through such a process (Eden & Ackermann, 1998). Whatever the data collection method, it is important that researchers are aware of the advantages and disadvantages of different data-collection options and use a method that suits their study's purpose.

Issues Related to Data Aggregation

If the researcher believes that the cognitive structure of various individuals can be aggregated to discern shared cognition and assessing shared cognitive structure is a key goal, then he/she needs to consider various options for data aggregation. Key questions to consider are (a) how to capture shared cognitive structure? And (b) what level of agreement is acceptable? Regarding the first question, Weick and Bourgon (1986) suggest three ways of computing shared maps based on individual maps: composite, assemblage, and average. These approaches were described earlier, so we do not repeat

them here. The decision on which approach to use largely depends on the research context.

Level of agreement on cause maps varies depending on the phenomenon studied and how well that phenomenon is shared, the managers investigated, and the precision of the mapping instruments. For example, the level of agreement would not be relevant if a company's annual reports are content analyzed to generate maps. If the level of agreement is to be examined, it is important to recognize that it is almost impossible to get 100% agreement among managers. The greater the number of individual maps used to compute the aggregate map and the more diverse these managers' views are on the problem or issue being investigated, the lower the level of agreement. If a research question involves the extent of sharing of cause maps, researchers could asses the extent of commonality in cause maps and how varying levels of commonality impact the quality and consistency of managerial decisions (Gnyawali, 1997).

Issues Related to Data Presentation and Drawing Inferences

Cause maps can become very complex if participants report all the linkages they see among a number of concepts. Some argue that this creates a tendency to show "everything related to everything else" and a tendency to place order on recollected events (Huff & Fletcher, 1990). One way to address this is to ask only for the most important relationships. Regarding the depiction of cause maps, Laukkanen (1998) suggests that researchers need to be careful in presenting the details in the map so that the reader is not overburdened with many complex maps. Researchers can get carried away with fancy presentations of maps and fail to provide clear insights based on the maps. Therefore, a key concern of the researcher should be the key message and insights generated from the study and conveyed by the maps.

A choice of the number of linkages to be reported may also depend on the extent to which similarities are expected or sought across individual maps and the extent to which the research question is focused on comparison of differences in perceived linkages. Jenkins (1998) identifies saliency and comparability as key issues in cause mapping. Saliency refers to "capturing the variables and relationships which accurately reflect the cognition of the individual," whereas comparability refers to "ensuring that there is sufficient commonality between the maps to make meaningful comparisons" (Jenkins, 1998, pp. 240–241). Researchers need to be aware of the saliency and comparability issues, and develop plans to balance them.

In the illustrative study described earlier, we addressed this balance by having semi-structured interviews with managers to understand the context and their views on the domain of study. A set of concepts were identified based on early interviews, and in the actual study the managers were allowed to delete any concept or add relevant concepts if anything was missing. Then, managers were asked to identify cause–effect links based on the final set of constructs they believed were important. Salience was improved because concepts were generated during early interviews before asking managers to complete the adjacency matrix, and validated later through written communication and conversations with managers. The research domain was very clearly defined (managers' view of their business's MO), so all the maps reflected managers' views of the same domain for their business. Comparability was enhanced because the 20 concepts (developed from the initial interviews) were common to everyone so the maps could be compared using these concepts and possible links among them.

Reliability relates to the question of how well can different researchers replicate the study. Reliability is arguably higher if the researcher uses standardized concepts in mapping so that the same concepts can be used in subsequent research. Similarly, more precise and detailed coding rules will make it easier for the future researchers to use the established rules in coding the data. Researchers need to clearly articulate the coding process and mechanisms (including the coding manual, training of the coders, how coding disagreements were resolved) so that future researchers could use the method and replicate the study. Reliability is also enhanced when researchers clearly articulate the domain of study, the nature of participants, and the process involved in eliciting the concepts.

Issues Related to the Use of Cause Mapping Software

The use of computer software for cause mapping is becoming increasingly popular in recent years. While software helps to systematize various aspects of cause mapping, it also makes the mapping process quite rigid because of the constraints of the software itself. We identify below a few advantages and disadvantages of using computer software.

Advantages of computer-assisted mapping

1. Complexity can be handled relatively easily
2. Possibly less time consuming

3. Faster visualization, which may trigger new thinking
4. Easier depiction of maps (maps can be edited and displayed with a few manipulations)
5. Easier to merge constructs
6. Easier to identify paths
7. Quantitative analysis of maps might be easier
8. Could be done in a distributed (asynchronous) environment through the web or group decision support system

Disadvantages of computer-assisted mapping

1. High standardization in the computer software may limit capturing of idiosyncrasies
2. Richness of information may be lost in the process of eliciting concepts and analysis of the maps if careful qualitative analysis is not done
3. Difficulty in administrating anytime, anywhere
4. Participant hesitation to engage in a computer-mediated process
5. Costs of computers and software relative to manual solicitation

We have identified two reasonably well-known software programs that have been used in cognitive mapping and briefly describe them below. They are Decision Explorer, formerly known as Graphics Cope (Cossettee, 2002; Eden et al., 1992) and CMAP2 (Laukkannen, 1994).

Decision Explorer

This program (available at www.banxia.com) can be used to generate concepts, prepare maps, and analyze the maps. Concepts are generated in a group brainstorming session and inputted in the computer. Links between concepts can be generated by using a mouse. A variety of analyses could be done using the software itself (it has a pull-down menu system and "analysis" is one of those menus). Decision Explorer can be used to identify concepts that have particular significance, for example, concepts with a high number of surrounding links ("ins," "outs" or both). Cluster analysis can be used to identify groups of concepts that are tightly linked together, which typically cover a particular area of the issue being mapped. One can display each cluster individually, print them for feedback or use them for discussion. Other possible analyses include centrality, domain, and loop. Since concepts used by Decision Explorer are idiosyncratic to each respondent group and standardization of concepts is not common, it is difficult to compare

cognitive maps at both content and structure levels. Comparison, however, is easier once some quantitative measures (such as centrality) are computed.

CMAP2

CMAP2 is primarily a text-based software (Laukkanen, 1994), but certainly can be used to make graphical presentations. But it is not as sophisticated graphically as Decision Explorer. While Decision Explorer was developed and primarily used to diagnose and design interventions, CMAP2 was developed primarily for research purposes. Cognitive maps can be compared once the natural language used by the respondents is translated into a system of standardized concepts/constructs. Some quantitative measures of various focal maps can be computed and different maps can be compared across such measures. See Laukkanen (1994) for a detailed description of CMAP2. Also, see Bood (1998, pp. 222–226) for a comparative discussion of CMAP2 and Decision Explorer.

PROMISING DIRECTIONS FOR FUTURE RESEARCH

As we noted at the beginning of this chapter, there is a relatively long history of cause mapping in organizational research (e.g., Axelrod, 1976), and scholars have made conscious efforts to use cause mapping in studying various topical areas in strategic management (e.g., Barr et al., 1992; Calori et al., 1994; Laukkanen, 1994; Stubbart & Ramaprasad, 1988). Strategic decision-making and organizational change are perhaps the most studied areas using cause mapping methods. In this section, we briefly outline some promising areas that could be further investigated using cause mapping. We do not attempt to be exhaustive in identifying the areas. Instead, our goal is to suggest some areas that are promising but have not received adequate attention from strategic management scholars.

Cause mapping can be used to develop a better understanding of the organizational knowledge, a very important topic in strategic management. From the cognitive perspective, one could examine how knowledge is formed, how it is represented, and how it affects organizational decisions and behaviors. In his review and categorization of organizational cognition research, Walsh (1995) identified the origin or development of knowledge structures, the representation of knowledge structures, and the use or consequences of knowledge structures as broad topic areas. Walsh suggested

that various factors influence the origin of shared knowledge structures in organizations and such knowledge structures importantly influence outcomes. As noted earlier, shared causal maps of managers can be viewed as organizational knowledge structures. Cause mapping studies could enable researchers to investigate how factors such as personal background characteristics or experiences of managers influence their cognitive structures, and the influences such structures have on key decisions of the managers. This kind of holistic examination will enrich our understanding of how cause maps develop, what consequences do various kinds of maps have on the organization, and possible interventions to encourage change.

One of the central premises of strategic management is that organizations need to change their strategies and structures (both proactively and reactively) as the environment changes. The perspective of organizational learning suggests that shared mental models in the organization need to reflect new understanding as the environment changes (Gnyawali & Stewart, 2003). If managers' cause maps influence their action priorities and decisions, then it is important to assess the extent and nature of changes in managers' cause maps over a period of time and how such changes led to changes in strategies and the structure of the organization. While some research has examined changes in cause maps (e.g., Barr et al., 1992; Calori et al., 1994; Weber & Manning, 2001), more work needs to be done in this area. It has been very informative to learn how company top executives' understanding of environmental influences and their effect on company performance change or lack of change over an extended time (Barr et al., 1992), but to what extent was this also a reflection of managers at other organizational levels? Why did these executives fail to interpret information and act on it? Were they unwilling to consider warnings from managers at lower levels? Does strategy change follow changes in mental models or do changes in mental models follow changes in strategy? Does this process differ by level in the organization or function? Do maps of boundary spanners change more rapidly than those of managers with little external contact? Possible extensions to studies of organizational change include but are not limited to considerations of interventions intended to build consensus within organizations and trust between organizations (e.g., Markoczy, 2001; Vangen & Huxham, 2003).

Yet another interesting area for research would be to investigate the relationship between cognitive structures (maps) and network structures. With the growing popularity of the perspectives of social network theory and potential effects of networks on organizational cognition, it would be fruitful to investigate these two areas using cause-mapping techniques. For

example, do managers who frequently interact with each other (high networking) have similar cognitive maps, although they are in different divisions or departments? If yes, would similarity of cognitive structure serve as a cause or effect of the similarity in network structure? Do managers involved in communities of practice inside and across firms have similar causal maps of domains of understanding? Do people join communities of practice because they already have similar beliefs? How do these managers' causal understandings change due to their membership in these support communities? At the inter-organizational level, one could examine whether firms having strong network ties (such as joint ventures) with each other will have similar cause maps of their industry and competitive environment? Would high level of director interlocks among firms (dense network structure) lead to similarity of cause maps of the Board of Directors and of the top management teams of the firms? Examination of these and related research questions would integrate two important areas of research on strategic management, and provide interesting insights that cannot be gained by using more traditional methods.

Recent advancements in information technology and artificial intelligence should also motivate strategy researchers to recognize the implications of cause mapping for knowledge management within complex organizations. Not only is it important that we understand what managers see as causally related, companies may need to capture that understanding so choices within the organization can be consistently made (e.g., expert systems) or tacit knowledge can be codified before mangers leave the company. For example, Beazley, Boenisch, and Harden (2002) suggested that in the United States 19% of the workforce holding executive, administrative, and managerial jobs will retire in the next 5 years. Many major corporations are recognizing the need to codify in some way the cognitive understandings of these managers (Casher & Lesser, 2003). We believe that causal mapping would be a useful technique to use in these knowledge management initiatives.

CONCLUSIONS

In this chapter, our primary objective was to provide method-related guidelines to researchers on the entire spectrum of issues involved in cause mapping and to encourage researchers to use causal mapping techniques in strategy research. We challenged strategists to move beyond simplified conceptualizations of strategic choice and decision-making in organizations and investigate the mental models that depict the cause and effect beliefs of

managers and their effects on organizational actions. We have taken readers through the cause mapping process incorporating eight basic steps into three stages discussing the "nuts and bolts" of cause mapping. We then provided an illustration of each of these steps and stages, and outlined "key issues to consider." Finally, we concluded with a discussion of some promising research directions. It is our hope that future researchers will devote more attention to cause mapping methods and better inform strategic management research and practice.

ACKNOWLEDGMENTS

We express our appreciation to the Marketing Science Institute for their financial support of this study and recognize Prakash Nedungadi for his contributions to the project leading to this manuscript. The first author also acknowledges the Pamplin Summer Research Grant provided to him by the R.B. Pamplin College of Business at Virginia Tech to support this line of research.

Figure 2 and Table 1 are reprinted from Tyler, B. B. and Gnyawali, D. R., (2002) *Journal of Product Innovation Management*, *19*, 259–276. Copyright with permission from Elsevier.

REFERENCES

Alba, J. W., & Hasher, L. (1983). Is memory semantic? *Psychological Bulletin*, *93*, 203–231.

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Axelrod, R. (1976). *The structure of decision*. New Jersey: Princeton University Press.

Barr, P. S., & Huff, A. S. (1997). Seeing isn't believing: Understanding diversity in the timing of strategic response. *Journal of Management Studies*, *34*(3), 337–370.

Barr, P. S., Simpert, J. L., & Huff, A. S. (1992). Cognitive change, strategic action, and organizational renewal. *Strategic Management Journal*, *13*(1), 15–36.

Beazley, H., Boenisch, J., & Harden, D. (2002). *Continuity management: Preserving corporate knowledge and productivity when employees leave*. New Jersey: Wiley.

Bood, R. (1998). Charting organizational learning: A comparison of multiple mapping techniques. In: C. Eden & J. C. Spender (Eds), *Managerial and organizational cognition* (pp. 210–230). Beverly Hills, CA: Sage publication.

Borgatti, S. P., Everett, M. G., & Freeman, L. C. (1999). *Ucinet 5 for windows: Software for social network analysis*. Natick, MA: Analytic Technologies.

Bougon, M., Weick, K., & Binkhorst, D. (1977). Cognition in organizations: An analysis of the Utrecht Jazz orchestra. *Administrative Science Quarterly*, *22*, 606–639.

Calori, R., Johnson, G., & Sarnin, P. (1994). CEOs' cognitive maps and the scope of the organization. *Strategic Management Journal*, *15*, 437–457.

Casher, A., & Lesser, E. (2003). *Gray matter matters: Preserving critical knowledge in the 21st century*. Somers, New York: IBM Business Consulting Services Publication.

Collins, A., & Loftus, E. (1975). A spreading activation theory of semantic processing. *Psychology Review*, *82*, 407–428.

Cossette, P. (2002). Analysing the thinking of F. W. Taylor using cognitive mapping. *Management Decisions*, *40*(1/2), 168–182.

Eden, C., & Ackermann, F. (1998). *Making strategy: The journey of strategic management*. London: Sage Publication.

Eden, C., Ackermann, F., & Cropper, S. (1992). The analysis of cause maps. *Journal of Management Studies*, *29*(3), 309–324.

Eden, C., & Spender, J. C. (1998). *Managerial and organizational cognition*. Beverly Hills, CA: Sage publication.

Fuglseth, A. M., & Gronhaug, K. (2002). Theory-driven construction and analysis of cause maps. *International Journal of Information Management*, *22*, 357–376.

Gatignon, H., & Xuereb, J. M. (1997). Strategic orientation and firm and new product performance. *Journal of Marketing Research*, *19*(February), 77–90.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of Grounded Theory: Strategies for qualitative research*. New York: Aldine De Gruyter.

Gnyawali, D.R. (1997). *Creation and utilization of organizational knowledge: An empirical study of the effects of organizational learning on strategic decision making*. Unpublished Ph.D. Dissertation, University of Pittsburgh.

Gnyawali, D. R., & Stewart, A. C. (2003). A contingency perspective on organizational learning: Integrating environmental context, organizational learning processes, and types of learning. *Management Learning*, *34*(1), 63–89.

Haray, F., Norman, R., & Cartwright, D. (1965). *Structural models: An introduction to the theory of directed graphs*. New York, NY: Wiley.

Henard, D. H., & Szymanski, D. M. (2001). Why are some new products more successful than others?: An analysis of cumulative evidence. *Journal of Marketing Research*, *38*, 365–375.

Hitt, M. A., & Tyler, B. B. (1991). Strategic decision models: Integrating different perspectives. *Strategic Management Journal*, *12*(5), 327–351.

Hodgkinson, G. P., Maule, A. J., & Brown, N. J. (2004). Causal cognitive mapping in the organizational strategy field: A comparison of alternative elicitation procedures. *Organizational Research Methods*, *7*(1), 3–26.

Huff, A. S. (1990a). *Mapping strategic thought*. New York, NY: Wiley.

Huff, A. S. (1990b). Mapping strategic thought. In: A. S. Huff (Ed.), *Mapping strategic thought* (pp. 11–49). New York: Wiley.

Huff, A. S., & Fletcher, K. E. (1990). Coding the association of concepts. In: A. S. Huff (Ed.), *Mapping strategic thought* (pp. 403–412). New York: Wiley.

Huff, A. S., & Schwenk, C. R. (1990). Bias and sensemaking in good times and bad. In: A. S. Huff (Ed.), *Mapping strategic thought* (pp. 89–108). New York: Wiley.

Jelinek, M., & Litterer, J. A. (1994). Toward a cognitive theory of organizations. *Advances in Managerial Cognition and Organizational Information Processing*, *5*, 3–48.

Jenkins, M. (1998). The theory and practice of comparing causal maps. In: C. Eden & J. C. Spender (Eds), *Managerial and organizational cognition* (pp. 231–249). Beverly Hills, CA: Sage publication.

Keller, K. L. (1993). Conceptualizing, measuring, and managing customer-based brand equity. *Journal of Marketing*, *57*, 1–22.

Langfield-Smith, K. M., & Wirth, A. (1992). Measuring differences between cognitive maps. *Journal of the Operational Research Society*, *43*(12), 1135–1150.

Laukkanen, M. (1992). *Comparative cause mapping of management cognition: A computer database method for natural data*. Helsinki: Helsinki School of Economics and Business Administration Publications.

Laukkanen, M. (1994). Comparative cause mapping of organizational cognition. *Organization Science*, *5*, 322–343.

Laukkanen, M. (1998). Conducting causal mapping research: Opportunities and challenges. In: C. Eden & J. C. Spender (Eds), *Managerial and organizational cognition* (pp. 168–191). Beverly Hills, CA: Sage publication.

Lord, R. G., & Foti, R. J. (1986). Schema theories, information processing, and organizational behavior. In: H. P. Sins & D. A. Gioia (Eds), *The thinking organization: Dynamics of organization social cognition* (pp. 20–48). San Francisco: Jossey-Bass.

Lyles, M. A., & Schwenk, C. R. (1992). Top management, strategy, and organizational knowledge structures. *Journal of Management Studies*, *29*, 155–174.

Miles, M. B., & Huberman, A. M. (1984). *Qualitative data analysis: A Sourcebook of new methods*. Newbury Park, CA: Sage Publications, Inc.

Markoczy, L. (1997). Measuring beliefs: Accept no substitute. *Academy of Management Journal*, *40*(5), 1228–1242.

Markoczy, L. (2001). Consensus formation during strategic change. *Strategic Management Journal*, *22*(11), 1013–1031.

Markoczy, L., & Goldberg, J. (1995). A method for eliciting and comparing causal maps. *Journal of Management*, *21*(2), 305–333.

Meindl, J. R., Stubbart, C., & Porac, J. F. (Eds) (1996). *Cognition within and between organizations*. Thousand Oaks, CA: Sage Publications.

Narayanan, V. K., & Fahey, L. (1990). Evolution of revealed causal maps during decline: A case study of admiral. In: A. S. Huff (Ed.), *Mapping strategic thought* (pp. 109–134). New York: Wiley.

Priem, R. L., & Harrison, D. A. (1994). Exploring strategic judgment: Methods for testing the assumptions of prescriptive contingency theories. *Strategic Management Journal*, *15*, 311–324.

Porac, J. F., & Thomas, H. (1989). Managerial thinking in business environments. *Journal of Management Studies*, *26*(4), 397–417.

Rosch, E. (1973). On the internal structure of perception and semantic categories. In: T. E. Moore (Ed.), *Cognitive development and the acquisition of language*. New York: Academic Press.

Stubbart, C. (1989). Managerial cognition: A missing link in strategic management research. *Journal of Management Studies*, *26*(4), 325–348.

Stubbart, C. I., & Ramaprasad, A. (1988). Probing two executives' schematic knowledge of the U.S. steel industry using cognitive maps. *Advances in Strategic Management*, *5*, 139–164.

Stubbart, C. I., & Ramaprasad, A. (1990). Comments on the empirical studies and recommendations for future research. In: A. S. Huff (Ed.), *Mapping strategic thought* (pp. 251–290). New York: Wiley.

Thomas, J. B., Sussman, S. W., & Henderson, J. C. (2001). Understanding "strategic learning": Linking organizational learning, knowledge management, and sensemaking. *Organization Science*, *12*(3), 331–345.

Toulmin, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*, 189–208.

Tyler, B. B., & Steensma, H. K. (1998). The effects of executives' experiences and perceptions on their assessment of potential technological alliances. *Strategic Management Journal*, *19*(10), 939–965.

Tyler, B. B., & Gnyawali, D. R. (2002). Mapping managers' market orientations regarding new product success. *Journal of Product Innovation Management*, *19*, 259–276.

Vangen, S., & Huxham, C. (2003). Nurturing collaborative relations: Building trust in inter-organizational collaboration. *The Journal of Applied Behavior Science*, *39*(1), 5–31.

Walsh, J. P. (1988). Selectivity and selective perception: An investigation of managers' belief structures and information processing. *Academy of Management Journal*, *31*, 873–896.

Walsh, J. P. (1995). Managerial and organizational cognition: Notes from a trip down memory lane. *Organization Science*, *6*(3), 280–321.

Weber, P. S., & Manning, M. R. (2001). Cause maps, sensemaking and planning organizational change. *The Journal of Applied Behavioral Science*, *37*(2), 227–251.

Weick, K. E. (1979). *The social psychology of organizing*. New York, NY: Random House.

Weick, K. E. (1985). Sources of order in underorganized systems: Themes in recent organizational theory. In: Y. S. Lincoln (Ed.), *Organizational theory and inquiry* (pp. 106–136). Beverly Hills, CA: Sage.

Weick, K. E. (1990). Cartographic myths in organizations. In: A. S. Huff (Ed.), *Mapping strategic thought* (pp. 1–10). New York: Wiley.

Weick, K. E., & Bourgon, M. G. (1986). Organizations as cognitive maps: Charting ways to success and failure. In: H. P. Sins & D. A. Gioia (Eds), *The thinking organization: Dynamics of organization social cognition* (pp. 102–135). San Francisco: Jossey-Bass.

Yin, R. K. (1994). *Case study research: Design and method* (2nd ed.). Thousand Oaks, CA: Sage Publishing, Inc.

THE DIMENSIONALITY OF ORGANIZATIONAL PERFORMANCE AND ITS IMPLICATIONS FOR STRATEGIC MANAGEMENT RESEARCH

James G. Combs, T. Russell Crook and Christopher L. Shook

ABSTRACT

Organizational performance is widely recognized as an important – if not the most important – construct in strategic management research. Researchers also agree that organizational performance is a multidimensional construct. However, the research implications of the construct's multidimensionality are less understood. In this chapter, we use a synthesis of previous attempts to describe the dimensions of performance and our own analysis of performance measurement in the Strategic Management Journal to build a conceptual model of organizational performance and its dimensions. Our model suggests that operational performance and organizational performance are distinct, and that organizational performance can be further dimensionalized into accounting returns, stock market, and growth measures. The model has implications for how future

Research Methodology in Strategy and Management
Research Methodology in Strategy and Management, Volume 2, 259–286

ISSN: 1479-8387/doi:10.1016/S1479-8387(05)02011-4

research might advance understanding about performance and how empirical studies should conceptualize and measure performance.

Organizational performance is an important, if not the most important, construct in strategic management research (Rumelt, Schendel, & Teece, 1994). Indeed, the focus on performance differentiates strategic management from other fields (Meyer, 1991); the *raison d'être* of strategic management research is to increase understanding about the determinants of organizational performance and explain how managers can create superior performance (Meyer, 1991). To advance, therefore, the strategic management field must cumulate knowledge regarding theories that help explain organizational performance and prescribe ways that managers can adjust strategies to improve organizational performance (Carlson & Hatfield, 2004; Rumelt et al., 1994).

A key challenge in explaining organizational performance and making valuable managerial prescriptions, however, is the significant two-way interrelationship between theory development and construct measurement (Venkatraman & Grant, 1986). The utility of a construct for theory development depends on a tradeoff between its definitional scope and parsimony (Bacharach, 1989). Parsimoniously defined constructs have limited utility because fewer hypotheses follow logically. Overly broad construct definitions, however, create measurement problems wherein competing measures of the same theoretical construct might have little relationship (Schwab, 1980). Although there is a large and growing body of strategic management research, little attention has been paid toward conceptualizing and measuring key constructs (Hitt, Boyd & Li, 2004). Consequently, researchers possess too much discretion when choosing among alternative measures, which has led to excessive variance across findings and limited the field's ability to accumulate knowledge (Boyd, Gove & Hitt, 2005).

The need for conceptual clarity regarding a construct's boundaries, dimensionality, and appropriate measures appears particularly important when the construct in question is central to an entire field of inquiry, such as organizational performance is for strategic management (Rumelt et al., 1994). If the mission of strategic management research is developing and testing theory explaining organizational performance, and this can only be accomplished when key constructs are well understood, then strategic management research can only fulfill its mission once there is a common understanding regarding how to conceptualize and measure organizational performance.

Schwab (1999) suggests that there are two main requirements to gain a common understanding of a construct. The first requirement is to provide a "dictionary-like statement" that explains a construct's domain (p. 32). To date, the domain of organizational performance has been defined very broadly as the social and economic outcomes resulting from the interplay among on organization's attributes, actions, and environment (Andrews, 1971; Hrebiniak, Joyce & Snow, 1989). Such a broad domain allows researchers to claim any social or economic outcome measure that can reasonably be tied to an organization or its environment, which reduces the utility of organizational performance as a construct (Bacharach, 1989).

A key second requirement to gaining a common understanding of a construct is measure validation (Schwab, 1999). Measure validation involves establishing content, convergent, and discriminant validity. Content validity is present when experts agree that measures fall within the construct's domain. Convergent validity is present when there is a high degree of agreement among two or more different measures of the same construct, and discriminant validity is present when measures of different constructs do not converge.

Unfortunately, evidence thus far suggests that the validity of competing measures of organizational performance is quite low (e.g., Rowe & Morrow, 1999). The strategic management literature is replete with different and frequently unrelated organizational performance measures (Maltz, Shenhar & Reilly, 2003; Starbuck, 2004; Venkatraman & Ramanujam, 1986). When multiple measures of the same construct are unrelated, the construct's validity is dubious. The measures might depict different dimensions of a broadly defined construct, but if the dimensions are unspecified and their boundaries unknown, there is no way for researchers to reconcile the seemingly conflicting findings that result naturally (Boyd et al., 2005; Hunter & Schmidt, 1990).

Despite the apparent need to improve how we conceptualize and measure organizational performance, little research attention has been devoted to advancing our understanding of the construct. We take a step forward by developing a model of organizational performance and its dimensions. The model is built on our synthesis of prior studies describing the dimensions of performance and our own analysis of performance measurement in the *Strategic Management Journal* (*SMJ*). Our central theme is that once researchers arrive at a consensus regarding the boundaries and dimensions of organizational performance, the dimensions should form the basis for accumulating evidence regarding theories of organizational performance. Instead of expecting an organizational action (e.g., strategy) or

environmental attributes (e.g., dynamism) to affect the overall organizational performance, theory can be developed explaining how different performance dimensions are affected by a given phenomenon. Based on our model, we offer advice regarding how strategic management researchers can design primary research that best helps the field achieve its mission to accumulate knowledge about the determinants of organizational performance.

EFFORTS TO UNDERSTAND ORGANIZATIONAL PERFORMANCE

Prior efforts to describe organizational performance can be traced back to organizational theorists' investigations of "organizational effectiveness" (Cameron & Whetten, 1981; Steers, 1975). Organizations were viewed as effective if they satisfied their stated goals (Lewin & Minton, 1986). But goals, such as profitability and growth or increased employee wages and lower prices, often conflict. Organizational effectiveness similarly encompasses the satisfaction of stakeholders with conflicting agendas. By including the satisfaction of competing goals and stakeholder claims within the construct's definition, organizational theorists found it difficult to empirically distinguish between effective and ineffective organizations, and little consensus emerged among researchers as to how to best measure effectiveness (Cameron, 1986; Chakravarthy, 1986).

Venkatraman and Ramanujam (1986) attempted to narrow the measurement domain for strategic management researchers with a model consisting of three concentric circles. The outer circle is organizational effectiveness, which they argued is too broad in scope to be practically applied in strategic management research. The middle circle is operational performance, which is represented by nonfinancial indicators of specific areas of organizational operations, such as product quality, innovation, or marketing outcomes. The innermost circle is financial performance, which includes measures relating to economic outcomes such as sales growth, accounting returns (e.g., return on investment (ROI)), and the stock market.

Because of the applied nature of strategic management research and the empirical difficulties with organizational effectiveness that stem from its inherently conflicting elements, Venkatraman and Ramanujam (1986) urged strategic management researchers to focus on the measurement domain identified by the inner two circles. However, they also recognized that even within the operational and financial performance circles there are numerous

operational (e.g., innovation or marketing) and financial (e.g., asset growth or accounting returns) performance outcomes, and measures of such outcomes often do not converge. Thus, even within these two measurement domains, substantial multidimensionality remains. Consequently, they suggested that researchers should either "explicitly test the dimensionality of their conception of business performance" or use "an a priori classification which recognizes the dimensionality issue" (p. 807).

Some researchers have taken Venkatraman and Ramanujam's advice to test the dimensionality of performance. For example, in a sample of new ventures, Florin, Lubatkin and Schulze (2003) examined the influence of human and social capital on both accounting returns and sales growth. They found support for the hypothesized effects of human and social capital on accounting returns, but not for sales growth. Robinson and McDougall (1998) similarly found that the degree of product differentiation influences accounting returns, but not the sales growth of industry participants.

However, Venkatraman & Ramanujam's (1986) suggestion that researchers use "an a priori classification which recognizes the dimensionality issue" (p. 807) has not been followed. Excluding studies that focus on organizational effectiveness, we found only five studies that attempted to build a classification of the underlying dimensions of organizational performance (e.g., Maltz et al., 2003; Murphy, Trailer & Hill, 1996; Rowe & Morrow, 1999; Tosi, Werner, Katz & Gomez-Mejia, 2000; Woo & Willard, 1983). One conducted a qualitative review (e.g., Murphy et al., 1996) and the other four used empirical analyses to derive related, but distinct performance dimensions. Table 1 highlights key similarities and differences across the studies.

Based on a review of the entrepreneurship literature, Murphy et al. (1996) identified four main performance dimensions: (1) efficiency (e.g., return on equity (ROE)), (2) growth (e.g., sales growth), (3) profitability (e.g., net income), and (4) size (e.g., net sales). They also suggested survival as an additional dimension. Of the four empirical analyses, three studies used factor analysis. Woo and Willard (1983) factor analyzed 14 performance measures using PIMS data and found four dimensions: (1) profitability (e.g., return on sales (ROS)), (2) relative market position (e.g., product quality vis-à-vis competitors), (3) change in profitability and cash flow (e.g., ROI variability), and (4) growth in sales and market share (e.g., market share gain). Similarly, Tosi et al. (2000) factor analyzed 30 performance measures using COMPUSTAT data to identify eight dimensions. Rowe and Morrow (1999) used confirmatory factor analysis to show that: (1) subjective (e.g., *Fortune* reputation surveys), (2) financial/accounting (e.g., return on assets

Table 1. Organizational Performance Studies.

Study	Dimensions Identified	Measures	Approximate Equivalent Dimension in Figure 1
Woo and Willard (1983)	Profitability	ROI, ROS, Cash flow to investment	Accounting returns
	Relative market position	Product quality, new product development, and costs vis-à-vis competitors, product R&D, process R&D	Marketing and sales, operations, and infrastructure outcomes
	Change in profitability and cash flow	Variation in ROI and cash flow to investment ratio	Accounting returns
	Growth in sales and market share	Revenue growth, market share, market share gain	Growth
Murphy, Trailer, and Hill (1996)	Efficiency	ROE, ROI	Accounting returns
	Liquidity	Quick ratio, current ratio	Firm (financial) capability, not a performance dimension
	Profit	EPS, net income	Accounting returns
	Size	Net sales, number of employees	Size, not a performance dimension
Rowe and Morrow (1999)	Subjective	*Fortune* reputation survey on management quality, financial soundness, value as long term investment, wise use of corporate assets	Operational effectiveness
	Financial (accounting)	ROA, ROI, cash flow over equity	Accounting returns
	Market	Sharpe, Treynor, Jensen's alpha/ unsystematic risk	Stock market

Tosi, Werner, Katz, and Gomez-Mejia (2000)	Absolute financial performance	Pre-tax profits, net income, stock price change	Accounting returns, size, and stock market
	Change in financial performance	Change in pre-tax profits, change in ROE, changes in net income	Profit growth
	Stock performance	EPS, 5 year average EPS, 5 year average EPS vs. industry average	Accounting returns/growth hybrid
	Return on equity – short term	ROE, ROE vs. industry average	Accounting returns
	Return on assets	ROA, ROA vs. industry average	Accounting returns
	Return on equity – long term	5 year average ROE, 5 year average ROE vs. industry average	Accounting returns
	Market returns	2 year average market return, market return	Stock market
	Internal performance indicators	Changes in working capital, market to book	Infrastructure outcomes and stock market
Maltz, Shenhar, and Reilly (2003)	Financial	Sales revenue, profit margin, revenue growth	Accounting returns and growth
	Market/customer	Customer satisfaction index, customer retention rates, service quality	Service outcomes
	Process	Time to market, quality of NPD and project management processes	Infrastructure and operations outcomes
	People development	Retention of top employees, quality of leadership development	Human resources outcomes
	Future	Depth of quality of strategic planning, anticipating/preparedness for environmental changes	Infrastructure outcomes

(ROA)), and (3) market (e.g., Sharpe ratio) performance measures were distinct, but still significantly related to an overall latent performance construct. Finally, Maltz et al. (2003) surveyed senior managers to derive five performance dimensions.

Overall, these studies agree that performance is a multidimensional construct, but the actual dimensions identified appear to depend on the analytical method (qualitative vs. quantitative), data source, and measures examined. Rowe and Morrow (1999) examined 10 measures and found 3 dimensions, Woo and Willard (1983) examined 14 measures and found 4 dimensions, and Tosi et al. (2000) examined 30 measures and found 8 dimensions. Thus, examining more measures appears to lead to the conclusion that there are more distinct dimensions, which further illustrates the low level of convergent validity among the wide variety of available alternatives.

The studies also combine different measures into similarly labeled dimensions. For example, Maltz et al. (2003) classify total sales, accounting returns, and sales growth under "financial," whereas Woo and Willard (1983) include accounting returns and sales growth, but not total sales, under "profitability," and Rowe and Morrow (1999) limit their "financial" dimension to just accounting returns. Thus, in each of these attempts to develop an a priori classification of organizational performance dimensions, there is not only disagreement regarding the number of dimensions, but also differences regarding the name of each dimension and which measure(s) belong to each dimension.

Nevertheless, some commonalities are discernable among the dimensions identified. First, each study points toward accounting returns (e.g., ROA) as all or part of one or more dimensions. Second, there is preliminary evidence that growth is a distinct dimension. Specifically, although two studies combine sales growth and accounting return measures, two others treat growth as a unique dimension – though one study uses sales growth whereas the other uses market share growth. Third, the two empirical studies that include stock market measures found that these are distinct from accounting returns and growth. Indeed, Tosi et al. (2000) described two separate stock market dimensions. Finally, two of the studies separate operational performance from organization-wide performance. Consistent with the notion that operational performance corresponds to firms' different operational activities, managers surveyed by Maltz et al. (2003) identified four distinct operational performance dimensions. Woo and Willard (1983) also identified a distinct operational performance dimension. The notion that operational performance should be considered outside the domain of organizational performance is also consistent with recent efforts to focus

resource-based theory on operational instead of organizational performance (Ray, Barney & Muhanna, 2004), and with calls for study of the relationship between operational performance and organizational performance (Priem & Butler, 2001).

Taken together, these studies point toward a broad consensus that performance is multidimensional and also begin to uncover the dimensions. There appears to be some agreement that accounting returns form at least one distinct performance dimension, and preliminary evidence that growth and stock market measures form at least two additional dimensions. Further, operational performance appears distinct from and outside the domain of organizational performance. Pulling together the findings of these initial efforts is an important first step. However, given that consensus regarding the dimensionality of key constructs is a prerequisite for theoretical development (Venkatraman & Grant, 1986), further efforts to increase understanding seem warranted. Indeed, more effort is needed to establish the dimensions of organizational performance. To this end, we offer an exploratory investigation of how organizational performance has been measured in extant strategic management research. Based on the information provided in the studies above, we also examined how the measures either converge with or discriminate from one another.

METHOD

We limited our analysis to articles published in the *SMJ* because of its reputation for publishing high-impact research and because we are confident that articles published in *SMJ* fall within the domain of strategic management (Shook, Ketchen, Cycyota & Crockett, 2003).

The analysis proceeded in three steps. First, we followed Steers' (1975) recommendation that in order to understand a construct, researchers must "consider how researchers have operationalized and measured the construct" (p. 546). Thus, we examined a random one half of the articles published in *SMJ* from its first issue in 1980 through October 2004 and recorded every performance measure used. Following Venkatraman and Ramanujam (1986), we coded a measure as a performance measure if it fell within the financial or operational circle of their model. We considered a measure operational if it reflected an outcome that could be tied to a specific value chain activity as described by Porter (1985), but did not reflect the interactive outcome of all value chain activities. Measures that depict outcomes attributable to the interaction among all value creation activities and the organization's environment were treated as organizational performance measures.

The second step of the analysis was to examine all studies published in *SMJ* from the first issue in 1980 through October 2004 and record all reported correlations between different organizational performance measures. We did this in order to examine how performance measures were related (i.e., converge or discriminate), and to determine if patterns emerged regarding relationships among measures. Whenever we found multiple studies that reported a correlation between the same two measures, we used meta-analysis to obtain an estimate of the overall weighted correlation between measures. For example, we found nine studies that reported a correlation between ROA and ROS. Following Hunter and Schmidt (1990), we calculated the sample-size weighted mean correlation between ROS and ROA for the nine studies. This procedure offers a greatly improved estimate of the actual population correlation. We focused on correlations among organizational measures because there were too few operational performance measures to conduct a meaningful analysis. Even among organizational measures, no study reported a correlation for many two-measure combinations, such as market-to-book value (MB) and operating margin (OM). The result was a partially complete correlation matrix that summarizes the known correlations to date, as reported in *SMJ*, between alternative organizational performance measures.

In the third step, we took the correlations from those measures where we had a complete correlation table and entered them into a confirmatory factor analysis. We did this to assess the dimensionality of organizational performance using the enhanced correlation estimates obtained from the meta-analyses. There were six measures – ROA, ROS, ROI, ROE, annual stockholder returns, and sales growth – where at least one study reported a correlation for each cell of the correlation table. The correlation table created from the meta-analysis of correlations among these six variables was entered into a confirmatory factor analysis using LISREL. The *n*-size was determined using the geometric mean of the total sample size upon which the meta-analytic correlations were derived. The geometric mean is a conservative measure of central tendency appropriate for cases where sample size varies dramatically (Hedges & Olkin, 1985).

RESULTS

Table 2 shows the variety of performance measures that have been used in *SMJ*. Overall, 56 different measures were used to depict operational or organizational performance. Of the 374 studies examined, 238 were

Table 2. Organizational and Operational Performance Measures.

	Frequency		Frequency
Organizational Performance Measures (367)			
Accounting returns	52%	*Growth*	17%
Return on assets	64	Sales	38
Return on sales	34	Profit	10
Return on equity	26	Market share	7
Return on investment	15	Employment	3
Operating margin	14	Growth scale[a]	3
Net income	13	Assets	2
Profit scale[a]	9	Earnings per share (EPS) growth	2
Combined accounting measures	5		65
Cash flow/assets	4	*Hybrids*	5%
Earnings per share	3	Growth/market share scale[a]	6
Net income/employees	2	Financial/growth scale[a]	5
Cash flow/sales	1	Stock price/earnings (P/E)	2
	190	Overall performance scale[a]	2
Stock market	11%	Cash flow/market value	1
Stock returns	15		16
Market to book value (Tobins Q)	15	*Survival*	6%
Jensen	5	Failure	21
Sharpe	2	Bankruptcy	1
Treynor	2		22
Security analyst assessments	1	*Other*	9%
	40	Market share	34
Operational Performance Measures (83)			
Marketing	10%	*Technology development*	11%
Sales/x[b]	7	Number of new products	5
Repeat business	1	IT performance scale[a]	1
	8	New product development time	1
Outbound logistics	5%	New product sales growth	1
Delivery time	2	Innovation scale[a]	1
Export performance scale[a]	1		9
Export sales	1	*Infrastructure*	2%
	4	Board effectiveness scale[a]	1
Operations	31%	Collaborative success scale[a]	1
Product quality scale[a]	10		2
Occupancy/load rate	4	*Hybrids*	39%
Costs/x[c]	3	Cumulative abnormal returns	26
Change in costs	3	Employee satisfaction[a]	6
Accident rate	2		32
Patents	2	*Human resources*	1%
Labor costs/x[c]	2	Employee turnover	1
	26		
		Service quality	1%
		Customer satisfaction	1

[a]Indicates survey measures.
[b]Indicates survey measures that depict the scale of marketing activities, such as firm marketing expenses.
[c]Indicates survey measures that depict the scale of production, such as sales in units or dollars.

empirical studies that measured performance a total of 450 times. The focus was on organizational performance 82% of the time, reflecting the field's interest in the relationships between internal firm activities and external environmental influences on performance (Andrews, 1971). Of the times organizational performance was measured, 52% examined accounting returns, 17% reflected growth over a period of time, 11% measured either stock market returns or value, market share was used 9% of the time, 6% depicted organizational survival, and 5% were hybrid measures that blended elements of accounting returns, growth, and stock market returns. Among the 83 instances operational performance examined, 31% were cumulative abnormal returns (CARs) to shareholders following specific events that can be tied to a single value chain activity. Wright, Ferris, Hiller and Kroll (1995), for example, contrasted shareholder reactions to winning a diversity program award with reactions to losing a discrimination lawsuit, and O'Shaugnessy and Flanagan (1998) investigated reactions to layoff announcements. Both examples reflect shareholder reactions to activities within the human resource value chain activity. If CARs are removed as a special case, 46% of the remaining instances focused on firm operations, 19% of these on product or service quality. In sum, Table 2 indicates that strategic management researchers have used a wide variety of performance measures, but the majority are organization-wide measures, and over half of these depict accounting returns.

Table 3 shows the partially complete correlation matrix from the meta-analysis of studies that report correlations between two or more organizational performance measures. We grouped the correlations into the three dimensions found in our review of prior studies of the dimensionality of performance (Table 1 (i.e., accounting returns, stock market, and growth)) – followed by measures that did not fall clearly into the major dimensions (i.e., market share, survival, and subjective measures of overall performance).[1]

The intragroup correlations among the accounting return measures were the strongest with the lowest correlation between ROE and OM at $r = 0.54$. In contrast, the highest correlation between any accounting return and any stock market measure is between ROA and market to book (MB) at $r = 0.32$. Although the relationship between accounting returns and profit growth is $r = 0.63$ for both ROA and ROE, the highest sales growth correlation is with OM at $r = 0.34$. Finally, accounting returns correlate between $r = 0.22$ (ROE) and $r = 0.40$ (ROI) with market share, close to zero (i.e., between $r = 0.06$ (ROA) and $r = 0.09$ (ROS)) with survival, and range between $r = 0.55$ (ROA) and $r = 0.76$ (ROS) with subjective survey

Table 3. Correlations among Organizational Performance Measures.

	ROA	ROS	ROE	ROI	OM	EPS	Stock	MB	Jensen	Sharpe	Treynor	Sales Growth	MS Growth	Profit Growth	Market Share	Survival	Subjective	P/E
Accounting returns																		
ROA	**1**	9/2237	10/3440	5/780	3/394	1/562	8/1150	8/1811	4/585			12/1752		2/150	7/2125	1/1588	4/315	1/80
ROS	0.87**	**1**	1/110	1/110	1/25		2/58	3/412	2/358			3/179				1/1588	1/42	
ROE	0.6**	0.6**	**1**	3/161	1/90		1/80	2/441				5/859		1/286	1/133			
ROI	0.83**	0.8**	0.73**	**1**	1/90		1/117					3/519			2/490			
OM	0.61**	0.84**	0.54**	0.83**	**1**							1/25						
EPS	0.8**					**1**												
Stock market																		
Stock	0.29**	0.19**	−0.11	0.24**			**1**					1/51		1/181				
MB	0.32**	0.23	0.1			0.57**		**1**	2/209			3/362			1/133			
Jensen	0.27**	0.19**						0.06*	**1**	1/200	1/200							
Sharpe									0.9**	**1**	1/200							
Treynor									0.87**	0.84**	**1**							
Growth																		
Sales Growth	0.32**	0.33**	0.16*	0.25	0.34		0.13	0.02				**1**	1/205	3/387	1/133	1/973	1/105	
MS Growth												0.8**	**1**					
Profit Growth	0.63**		0.63**				0.89**					0.53**		**1**	1/205			
Hybrids and Other Measures																		
Market Share	0.37**		0.22**	0.4*				0.24**				0.01		0.56**	**1**			
Survival	0.06*	0.09**										−0.01				**1**		
Subjective	0.55**	0.76**										0.57**					**1**	
P/E	0.18																	**1**

Note: Correlations are below the diagonal and the number of studies/total sample size are above the diagonal.
ROA = Return on Assets, ROS = Return on Sales, ROE = Return on Equity, ROI = Return on Investment, OM = Operating Margin, EPS = Earning per share, Stock = Stock Returns, MB = Market-to-Book Value, Jensen = Jensen's Alpha, Sharpe = Sharpe Ratio, Treynor = Treynor Index, Sales Growth = Sales Growth, MS Growth = Market Share Growth, Profit Growth = Profit Growth, Subjective = Subjective Survey Measure and P/E = Price to Earnings Ratio.
$^{*}p<0.05$.
$^{**}p<0.01$.

measures of overall performance. Overall, measures of accounting returns show evidence of convergent validity and, with the exception of profit growth, discriminant validity from other performance dimensions. Also, accounting returns appear to figure prominently in managers' survey evaluations of overall performance.

One study showed high convergent validity among Jensen's alpha, the Treynor index, and the Sharpe ratio,[2] but no study relates these measures to actual shareholder returns, which is a more common stock market measure in strategic management research (38%, Table 2). Also, although MB is a fairly common stock market measure (38%, Table 2), there is evidence it is unrelated to Jensen's alpha ($r = 0.06$). Thus, evidence of convergent validity among alternative stock market measures is mixed. Finally, although the evidence suggests discriminant validity between stock market measures and accounting returns, based on two separate studies, stock measures appear strongly related to profit growth ($r = 0.89$) but not sales growth ($r = 0.13$). No correlations were available between stock market measures and market share, survival, or subjective measures of overall performance. Overall, stock market measures discriminate from accounting returns and sales growth, but not profit growth, and evidence of convergent validity is mixed.

Among growth measures, sales growth was highly related to market share growth in one study ($r = 0.80$), but sales growth and profit growth were only related at $r = 0.53$. Further, the few studies that related profit growth to other performance dimensions report that it is related to accounting returns (i.e., $r = 0.63$ for both ROA and ROE) and annual stock returns ($r = 0.89$), whereas the most common growth measure, sales, is not. The only other performance measure related to sales growth appears to be managers' subjective assessments of overall performance at $r = 0.57$. Overall, growth measures demonstrate good convergent and discriminant validity only once profit growth is removed. The data suggest that whereas profit growth is only moderately related to any other growth measure, it is also moderately related to accounting returns and strongly related to a stock market measure.

As for the last three measures in Table 3, market share is a measure of a firm's relative size within its industry (Lenz, 1981). Thus, it is arguably not a performance measure (Buzzell, Gale & Sultan, 1975). It was included, however, because of its frequent use as such. Not surprisingly, given the effect of size on profits via economies of scale (Buzzell et al., 1975), market share shows significant correlations with accounting returns (from $r = 0.22$ for ROE to $r = 0.40$ for ROI). Only one study reported correlations between survival and accounting returns and another reported a correlation with

sales growth; each was close to zero. Thus, preliminary evidence suggests survival is independent of the other dimension.

In the third analysis, we entered the correlations of the six performance measures for which we had complete correlations (i.e., ROA, ROS, ROE, ROI, stock returns, and sales growth) into a confirmatory factor analysis. Based on the literature review of prior studies that dimensionalize performance (Table 1) and our examination of the correlations (Table 3), we initially predicted that ROA, ROS, ROE, and ROI would comprise one factor, and stock market returns and sales growth would load on two additional factors. We used LISREL to conduct the confirmatory factor analysis and assessed the three aspects of model fit (i.e., absolute, incremental, and parsimonious; Hair, Anderson, Tatham & Black, 1998) using four indices: Chi-square, goodness of fit index (GFI), adjusted goodness of fit index (AGFI), and normed fit index (NFI). Despite adjusting for sensitivity to sample size, the chi-square indicated poor model fit ($\chi^2 = 79.93$, df $= 8$, $p < 0.00$). The GFI (0.88) and NFI (0.88) were close to the recommended level of 0.90, but the AGFI (0.69) was not. We then tested a model that separated ROE from ROA, ROS, and ROI, and treated ROE as a distinct factor. This model was also not supported ($\chi^2 = 52.97$, df $= 6$, $p < 0.00$, AGFI $= 0.71$). We, therefore, removed ROE from the analysis because of its low correlation with the other accounting return measures and because its denominator is a measure of financial leverage, which is easily manipulated by managers. The model excluding ROE had good fit ($\chi^2 = 9.53$, df $= 4$, $p < 0.05$, GFI $= 0.98$, NFI $= 0.93$, and AGFI $= 0.98$) suggesting that, excluding ROE, accounting returns, shareholder returns, and sales growth reflect three distinct performance dimensions. We build on these findings to draw a model of performance dimensionality.

TOWARD A MODEL OF PERFORMANCE DIMENSIONALITY

Recent theoretical developments (e.g., Ray et al., 2004) and our review of the studies that previously attempted to describe the dimensions of performance suggested that operational performance is related to, but outside the conceptual domain of organizational performance (i.e., organization-wide performance). Further, at a minimum, accounting returns, growth, and the stock market tap different organizational performance dimensions. Our analysis of performance measurement in *SMJ* (Table 3) also shows evidence

of convergent validity among accounting returns and, to a lesser extent, among stock market measures. But, in both the review of studies and the analysis of performance measures in *SMJ*, the level of convergent validity depends on the measure. With the exception of profit growth, accounting return, stock market, and growth measures show discriminant validity. Although the number of measures included is small, our confirmatory factor analysis also points to these as empirically distinct dimensions. Thus, we conclude that accounting returns, the stock market, and growth represent three related but distinct dimensions of organizational performance. Although there might be other dimensions, such as survival, these three have the strongest empirical case that within-dimension measures converge and that between-dimension measures discriminate. Fig. 1 depicts these relationships and the logic supporting the model is explained below.

Operational Performance

In contrast to its earlier treatment as a dimension of a broadly defined organizational performance (Venkatraman & Ramanujam, 1986), recent theory (Ray et al., 2004) and the studies of performance dimensionality (Table 1) indicate that operational performance is a separate construct from

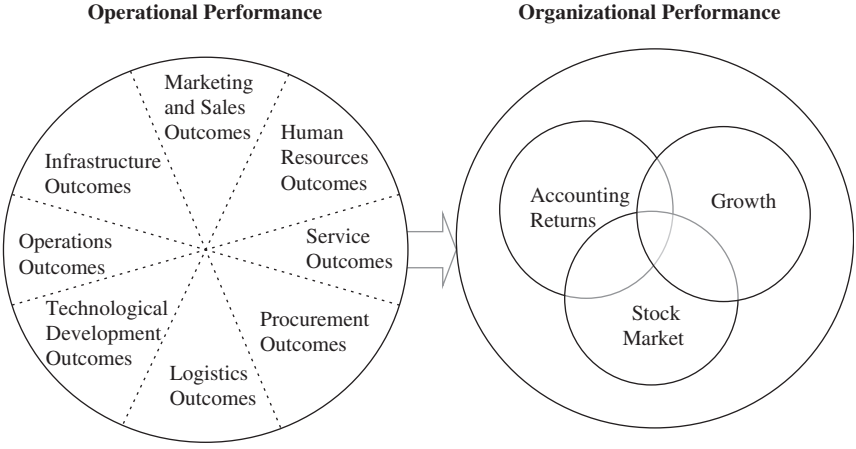

Fig. 1. Operational and Organizational Performance Dimensions. Note: Operational Performance Outcomes are Based on Porter's (1985) Value Chain Model.

organizational performance. Further, operational performance is itself multidimensional (Maltz et al., 2003; Porter, 1985). However, these studies do not suggest how operational performance should be viewed vis-à-vis organizational performance. Because organizational performance is affected, in part, by the sum of the firm's operational performance across many different value chain activities (Porter, 1985; Ray et al., 2004), we conclude that operational performance is best viewed as an antecedent to organizational performance. For example, when strategic human resource management researchers look at the relationship between high performance work practices (HPWPs) and performance, they view HPWPs as first affecting human resources (HR) outcomes such as productivity and employee turnover, which in turn benefit organization-wide performance (Dyer & Reeves, 1995; Huselid, 1995). In other words, operational performance mediates the relationship between internal activities (e.g., strategies, resources, and capabilities) and organizational performance (Ray et al., 2004). Acknowledging that operational performance outcomes are logical antecedents of organizational performance should empower researchers to explore how operational outcomes mediate relationships between strategy, resources, and capabilities and organizational performance (Priem & Butler, 2001; Ray et al., 2004).

As shown in Fig. 1, the domain of operational performance is still very broad and multidimensional. In Porter's (1985) description of value chains, he points out that there are many different intraorganizational activities that help bring a product or service to customers. Outcomes from these different activities, as shown in Table 2, have been used to measure operational performance. However, as Ray et al. (2004) point out, outcomes from different processes and activities within the firm should not be treated as highly related components of a unidimensional operational performance construct. In the context of resource-based theory, they argue that different processes and activities within firms have different effects on organizational performance.

These different processes and activities and their outcomes might have competing impacts on organizational performance. An organization that has a competitive advantage in customer service, for example, might have a disadvantage in production that negates any visible effects of the customer service advantage on organizational performance. The important point is that the outcomes of different activities are not necessarily related to one another. Thus, while it is convenient to use the overarching term operational performance, it is necessary to treat outcomes from different activities as independent dimensions and perhaps more importantly, to keep these dimensions separate from organizational performance.

Organizational Performance

Fig. 1 is drawn to depict the three related, yet distinct, dimensions of organizational performance suggested by our review of the studies that dimensionalize performance (Table 1) and in our own analysis of *SMJ* (i.e., accounting returns, the stock market, and growth). The first reason the dimensions overlap in Fig. 1 is the fact that they are both theoretically and statistically related (e.g., Fryxell & Barton, 1990; Rowe & Morrow, 1999). For example, a low-cost strategy might increase accounting returns by lowering costs and increase growth by attracting customers to lower prices (Campbell-Hunt, 2000). Similarly, growth can improve economies of scale, which makes firms more efficient and profitable (Buzzell et al., 1975; Makadok, 2000).

Although the dimensions are related, they also have key differences. Growth, for example, does not necessarily lead to accounting returns; firms might sacrifice profits to increase growth. Firms might similarly forgo growth for profit, particularly when overall industry growth is stagnant (Miles & Snow, 1978). The result is that each performance dimension is positively correlated, though not strongly enough to be considered alternative measures of a single one-dimensional construct.

A second reason the dimensions are shown as overlapping is because some measures are hybrids that capture elements of multiple dimensions. For example, our analysis revealed that surveys asking managers to rate their overall performance show strong correlations with both accounting returns and growth. Scales combining questions about growth and profitability have acceptable reliability (e.g., Capron, 1999; Lane, Salk & Lyles, 2001), which suggests that managers are able to consider accounting returns and growth simultaneously when assessing organizational performance. Other hybrids, such as the (stock) price to (accounting) earnings ratio (i.e., P/E), depict the organization's relative performance on two dimensions and thus might not relate strongly to either (e.g., for P/E-ROA, $r = 0.18$).

IMPLICATIONS FOR RESEARCH IN STRATEGIC MANAGEMENT

Although the model presented in Fig. 1 is derived from our examination of prior investigations of performance dimensionality and our analysis of the reported correlations among performance measures, we depart from prior literature in two important respects. First, although Venkatraman and

Ramanujam (1986) depicted organization-wide financial measures as a subset of operational performance, we view operational and organizational performance as conceptually distinct. Second, we identified three interrelated, yet distinct, dimensions of organizational performance (i.e., accounting returns, growth, and the stock market). These departures from prior literature have important implications in two areas: (1) they point toward important avenues for future inquiry to increase our understanding of performance and (2) they have practical implications for how to design studies where performance is the dependent variable.

Future Research Directions for Understanding Performance

From our examination of prior efforts to understand the dimensionality of performance we concluded that (1) operational performance is both different from and an antecedent to organizational performance and (2) operational performance has many dimensions. These conclusions are consistent with recent theoretical advances suggesting that organizational and operational performance are different and the relationship between them is affected differently for different operational activities (Ray et al., 2004). Other theory suggests that powerful stakeholders might siphon away some of the fruits of high operational performance and thus intervene between operational and organizational performance (Coff, 1999). Overall, researchers interested in further clarifying the dimensionality of performance need to begin by separating operational performance from organizational performance.

Porter's (1985) value chain describes the several known dimensions of operational performance, but research is needed to explain how these dimensions interact. For example, high productivity is an important manufacturing outcome and low turnover is an important human resource outcome. However, some organizations achieve productivity through a control approach that works employees hard and results in high turnover, whereas others use an empowerment approach that reduces turnover (Guthrie, 2001). Clearly, there are several causal and theoretically interesting relationships among the dimensions of operational performance that are potentially lost if researchers treat operational performance as a one-dimensional construct.

Our study found preliminary evidence in support of three interrelated dimensions of organizational performance – accounting returns, growth, and the stock market. This result points to at least two important questions:

Are there other dimensions? And how closely are the dimensions related? We initially defined organizational performance as the social and economic outcomes resulting from the interplay between an organization's attributes, actions, and environment (Andrews, 1971; Hrebiniak et al., 1989). Even after moving operational performance outside the construct domain, however, organizational performance is still broad. There are many economic and social outcomes, such as impacts on the natural environment and local communities, which lay beyond the three dimensions identified. Future research should focus on identifying and classifying these other outcomes. One potentially fruitful avenue might be to use managers as informants. Because managers and researchers often view phenomena differently (Reger & Huff, 1993), gaining practitioner insights might help identify additional dimensions.

One outcome that deserves attention is survival. Survival is often used as the ultimate measure of organizational performance (e.g., Fischer & Pollock, 2004). Murphy et al. (1996) argued that it is an independent dimension, and the three studies in our analysis that correlated survival to accounting returns and growth showed small or negative relationships, which also suggest that survival is a distinct dimension. However, other studies suggest that survival is preceded by organizational performance rather than being a dimension of organizational performance (e.g., Hudson, 1986). In short, additional research is needed that analyzes whether survival is a distinct dimension of organizational performance or an entirely different construct.

We were limited in our analyses by the number of studies reporting correlations among organizational performance measures. Thus, more work needs to be done to establish convergent and discriminant validity. One avenue would be to expand our analysis beyond *SMJ*. Another approach would be to increase the number of studies that test convergent and discriminant validity. The ultimate goal should be a complete Table 3 wherein each correlation is based on many studies.

The fruits of such efforts will be knowledge regarding which measures best depict each dimension and which measures should be avoided. In the stock market dimension, for example, Jensen's alpha, the Sharpe ratio, and the Treynor index are all measures of stock return, but their relationship with overall stock returns is still to be examined. Further, the two studies that relate Jensen's alpha, a measure of stock return, with the market-to-book ratio, a measure of stock value, only show a correlation of $r = 0.06$. Are stock returns and stock value empirically different? If so, use of both might lead to conflicting results. Similarly, more research is needed on profit

growth. Profit growth in our data has moderate to strong correlations with key measures in all three dimensions. Does that make profit growth a good universal measure? Future researchers will be better equipped to design fruitful studies if they could draw upon a body of research describing which measures are valid and reliable indicators of each dimension and which are not.

Practical Implications for Designing Studies

Our findings offer practical implications for strategy researchers involved in testing theory that explains organizational performance. Specifically, studies need to account for the relationship between operational and organizational performance, the multidimensional nature of organizational performance, and the interrelationships among the dimensions. Key implications that arise by accounting for these elements are depicted in Table 4.

Recognizing operational performance as distinct from organizational performance opens up the potential for strategy researchers to develop much richer theory. Studies are needed to investigate the relationship between operational and organizational performance, as well as how phenomena relate differently to each. For example, we might expect external environmental forces to more directly affect organizational performance than internal activities because the latter is mediated by operational performance (Ray et al., 2004). Indeed, studies are needed to test the extent to which operational performance mediates the relationship between internal activities, such as HPWPs and organizational performance (e.g., Huselid, 1995). Failure to explain how internal capabilities affect organizational performance is a central criticism of resource-based theory (Priem & Butler, 2001). We also need a theory describing factors that moderate the relationship between operational and organizational performance. Powerful stakeholders provide one example (Coff, 1999). More generally, recognizing the differences between operational and organizational performance should allow researchers to build richer theory and design more fine-grained empirical tests.

If operational and organizational performance are distinct, it follows that researchers should avoid measures that capture elements of both. For example, the numerator in ROE is derived from the accounting returns dimension of organizational performance, but the denominator reflects the firm's capital structure, which is an outcome of infrastructure (i.e., finance) in the value chain. This explains, at least in part, why ROE did not load

Table 4. Suggestions for Strategy Researchers.

Implications of Separating Operational from Organizational Performance

(1) Develop theory and empirical tests to learn which:
- Factors impact operational performance more than organizational performance and vice versa.
- Types of operational performance have the most impact on organizational performance.
- Factors moderate the relationship between them.

(2) Avoid measures that are composites of operational and organizational performance.

Implications of the Distinct Dimensions of Organizational Performance

(1) Select performance measures by matching the performance dimension with the underlying theory.
(2) Select and validate measures separately within each dimension of interest.
(3) Collect measures from multiple dimensions, but:
- do not necessarily expect convergence,
- use dimensionality to test the limits of theory, and
- to build separate bodies of knowledge around each dimension.

Implications of the Interrelated Nature of the Dimensions

(1) Establish the validity of selected measures by either:
- testing convergent and discriminant validity, or
- "sticking to" measures previously validated for the dimension of interest.

(2) Avoid "hybrid" measures in the overlapping areas of the performance dimensions.
(3) Design perceptual scales that cue respondents to the performance dimension of interest.

adequately in our confirmatory factor analysis. Conceptually, earnings per share (EPS) suffers from the same weakness.

Recognizing that organizational performance comprises at least three distinct dimensions also has implications for study design. One implication is that researchers should circumscribe theory and empirical tests to the appropriate dimension or dimensions. For example, transaction cost theory predicts that firms organize productive activity internally or using hybrid organizational forms (e.g., joint ventures) when the transaction costs (e.g., negotiating, monitoring) of using markets are high (Williamson, 1975). Doing so is efficient and thus improves performance (Poppo & Zenger, 1998). This theory, like others rooted in economics (e.g., resource-based theory), is focused specifically on efficiency, and thus should have its most profound impact on accounting returns. Resource dependence theory, in contrast, is focused on strategies for gaining access to resources, and thus might be expected to relate strategy to growth (Pfeffer & Salancik, 1978). Similarly, if the theory suggests that shareholders should care about

something, such as having an independent board of directors, perhaps a stock market measure is appropriate (e.g., Shivdasani & Yermack, 1999), even though we might not expect shareholders' concerns to translate into accounting profits (cf. Dalton, Daily, Ellstrand & Johnson, 1998).

A related implication is that researchers should not expect results from different tests of the same hypothesis to converge if the measures depict different dimensions. The value of triangulation has been widely accepted in the organization sciences (Jick, 1979). Although we encourage researchers to triangulate within dimensions by collecting multiple measures from multiple sources (e.g., secondary and primary; Venkatraman & Ramanujam, 1986), we are less sanguine about the notion that results based on different dimensions should converge. Acquiring another firm, for example, generally increases sales growth (Capron, 1999), but an acquisition's effect on shareholder returns depends on factors such as the relatedness of the target (Palich, Cardinal & Miller, 2000). Franchising similarly accelerates growth, but has an ambiguous effect on accounting returns (Combs, Michael & Castrogiovanni, 2004). We encourage researchers to collect measures from different dimensions for the purpose of testing the boundaries of the theory (Bacharach, 1989) and building a unique body of knowledge around each dimension, but not to triangulate on the determinants of organizational performance as if it were a one-dimensional construct.

Finally, the interrelated nature of the dimensions has important implications for the selection of measures. Confidence in a study's findings relies on a foundation of measures with known and reliable properties. Measures should be justified based on their appropriateness for the research setting and their validity as established in the literature. If multiple measures are used, their reliability, and convergent and discriminant validity must be empirically assessed. Unfortunately, too many studies justify combining disparate performance measures without gauging their reliability or validity, presumably on the grounds that performance is multidimensional and thus convergence is not necessarily expected (e.g., Anderson, Forsgren & Holm, 2002; Qian & Li, 2003). However, this practice weakens reliability and thus the likelihood of finding support for hypotheses (Hunter & Schmidt, 1990). Further, if a relationship is found, there is no way for the researcher to know which performance dimension is driving the result.

One approach to increased validity is to select measures that prior investigations have already validated as tapping the performance dimension of interest. Our study offers preliminary evidence that ROA, ROI, ROS are reliable measures of accounting returns, and that Jensen's alpha, the Sharpe ratio, and the Treynor index are all related stock market measures.

Weinzimmer, Nystrom, and Freeman (1998) offer advice for measuring growth. Researchers should also avoid "hybrid" measures that combine aspects from multiple dimensions. Measures such as the (stock) price to (accounting) earnings (i.e., P/E) ratio capture elements of two or more dimensions. Survey scales that combine items about different dimensions should also be avoided. Although such scales can be designed with adequate reliability (e.g., Wiklund & Shepherd, 2003), their content validity is questionable because it is difficult for researchers to assess the degree to which different performance dimensions are affected by the phenomena of interest. Thus, survey scales should be designed to recognize the dimensionality of organizational performance.

CONCLUSION

Although organizational performance is perhaps the most important construct in strategic management research, the nature of its definitional boundaries and dimensionality have been poorly understood. Based on a review of previous attempts to understand the dimensionality of organizational performance and our own analyses of performance measurement in the *SMJ*, we built a conceptual model of organizational performance. Our model has implications for the way organizational performance is conceptualized vis-à-vis operational performance, and for the way theory and empirical tests should be circumscribed to address the dimensionality of organizational performance. It is our hope that others will take up the challenge of advancing understanding of organizational performance, its dimensions, and its proper measurement. We also hope researchers will follow the practical implications of our model. Without such efforts, we confront increased risk of floundering in a sea of measure-driven conflicts among reported findings (Bacharach, 1989; Hunter & Schmidt, 1990). Alternatively, by understanding the dimensions of our central construct and building this knowledge into our theories and empirical tests, the field of strategic management is more likely to fulfill its mission to generate cumulative knowledge about the determinants of organizational performance.

NOTES

1. Although significance tests are reported in Table 3, our discussion focuses on the size of the relationship because it is size, not significance, that determines whether two measures depict the same construct (Schwab, 1999).

2. All three measures reflect the stock market return of a stock over some period of time relative to the stock's risk. The Sharpe ratio is $R_i - RFR_i/\sigma_i$ where R_i is the stock's return, RFR_i is the risk-free rate on U.S. Treasury Bills, and σ_i is the stock's price variance. Jensen's alpha is α_i from the regression equation $R_i = \alpha_i + \beta_i(R_{\mathrm{m}})$ where R_i is the stock's return and R_{m} is the market's overall return. The Treynor index is $R_i - RFR_i/\beta_i$ where β_i is the stock's price variance relative to the market variance as calculated in the same regression equation used to calculate Jensen's alpha.

ACKNOWLEDGMENT

We would like to acknowledge the assistance of Dilene Crockett and Cynthia Cycyota.

REFERENCES

Anderson, U., Forsgren, M., & Holm, U. (2002). The strategic impact of external networks: Subsidiary performance and competence development in the multinational corporation. *Strategic Management Journal*, *23*(1), 979–998.

Andrews, K. (1971). *The concept of corporate strategy*. Homewood, IL: H. Dow Jones-Irwin.

Bacharach, S. B. (1989). Organizational theories: Some criteria for evaluation. *Academy of Management Review*, *14*(4), 496–515.

Boyd, B., Gove, S., & Hitt, M. (2005). Construct measurement in strategic management research: Reality or illusion? *Strategic Management Journal*, *26*(3), 239–257.

Buzzell, R., Gale, B., & Sultan, R. (1975). Market share: A key to profitability. *Harvard Business Review*, *53*(1), 97–106.

Cameron, K. (1986). Effectiveness as paradox: Consensus and conflict in conceptions of organizational effectiveness. *Management Science*, *32*(5), 539–553.

Cameron, K., & Whetten, D. (1981). Perceptions of organizational effectiveness over organizational life cycles. *Administrative Science Quarterly*, *26*, 525–544.

Campbell-Hunt, C. (2000). What have we learned about generic competitive strategy: A meta-analysis. *Strategic Management Journal*, *21*(2), 127–154.

Capron, L. (1999). The long-term performance of horizontal acquisitions. *Strategic Management Journal*, *20*(11), 987–1018.

Carlson, K., & Hatfield, D. (2004). Strategic management research and the cumulative knowledge perspective. In: D. J. Ketchen & D. D. Bergh (Eds), *Research methodology in strategy and management* (pp. 273–301). San Diego, CA: Elsevier.

Chakravarthy, B. S. (1986). Measuring strategic performance. *Strategic Management Journal*, *7*, 437–458.

Coff, R. W. (1999). When competitive advantage doesn't lead to performance: The resource-based view and stakeholder bargaining power. *Organization Science*, *10*(2), 119–133.

Combs, J. G., Michael, S. C., & Castrogiovanni, G. J. (2004). Franchising: A review and avenues to greater theoretical diversity. *Journal of Management*, *30*(6), 907–931.

Dalton, D. R., Daily, C. M., Ellstrand, A. E., & Johnson, J. L. (1998). Meta-analytic reviews of board composition, leadership structure, and financial performance. *Strategic Management Journal*, *19*(3), 269–290.

Dyer, L., & Reeves, T. (1995). Human resource strategies and firm performance: What do we know and where do we need to go? *International Journal of Human Resource Management*, *6*, 656–670.

Fischer, H. M., & Pollock, T. G. (2004). Resetting the clock, sociopolitical transformational shields and IPO firm survival. *Academy of Management Journal*, *47*, 463–481.

Florin, J., Lubatkin, M., & Schulze, W. (2003). A social capital model of high-growth ventures. *Academy of Management Journal*, *46*(3), 374–384.

Fryxell, G., & Barton, S. (1990). Temporal and contextual change in the measurement structure of financial performance: Implications for strategy research. *Journal of Management*, *16*(3), 553–569.

Guthrie, J. P. (2001). High-involvement work practices, turnover, and productivity: Evidence from New Zealand. *Academy of Management Journal*, *44*, 180–190.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis* (5th ed.). Upper Saddle River, NJ: Prentice-Hall.

Hedges, L., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Hitt, M., Boyd, B., & Li, D. (2004). The state of strategic management research and a vision of the future. In: D. J. Ketchen & D. D. Bergh (Eds), *Research methodology in strategy and management* (pp. 1–31). San Diego, CA: Elsevier.

Hrebiniak, L., Joyce, W., & Snow, C. (1989). Strategy, structure, and performance. In: C. C. Snow (Ed.), *Strategy, organization design, and human resource management* (pp. 3–54). Greenwich, CT: JAI Press.

Hudson, J. (1986). An analysis of company liquidations. *Applied Economics*, *18*, 219–235.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis*. Newbury Park, CA: Sage.

Huselid, M. A. (1995). The impact of human resource management practices on turnover, productivity, and corporate financial performance. *Academy of Management Journal*, *38*, 635–672.

Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative Science Quarterly*, *24*, 602–611.

Lane, P., Salk, J., & Lyles, M. (2001). Absorptive capacity, learning, and performance in international joint ventures. *Strategic Management Journal*, *22*(12), 1139–1161.

Lenz, R. (1981). 'Determinants' of organizational performance: An interdisciplinary review. *Strategic Management Journal*, *2*, 131–154.

Lewin, A., & Minton, J. (1986). Determining organizational effectiveness: Another look, and an agenda for research. *Management Science*, *32*(5), 514–538.

Makadok, R. (2000). Interfirm differences in scale economies and the evolution of market shares. *Strategic Management Journal*, *20*(10), 935–952.

Maltz, A., Shenhar, A., & Reilly, R. (2003). Beyond the balanced scorecard: Refining the search for organizational success measures. *Long Range Planning*, *36*, 187–204.

Meyer, A. D. (1991). What is strategy's distinctive competence? *Journal of Management*, *17*(4), 821–833.

Miles, R., & Snow, C. (1978). *Organizational strategy, structure, and process*. New York: McGraw Hill.

Murphy, G., Trailer, J., & Hill, R. (1996). Measuring performance in entrepreneurship research. *Journal of Business Research*, *36*, 15–23.

O'Shaugnessy, K. C., & Flanagan, D. J. (1998). Determinants of layoff announcements following MandAs: An empirical investigation. *Strategic Management Journal*, *19*, 989–1000.

Palich, L., Cardinal, L., & Miller, C. (2000). Curvilinearity in the diversification-performance linkage: An examination of over three decades of research. *Strategic Management Journal*, *21*(2), 155–174.

Pfeffer, J., & Salancik, G. (1978). *The external control of organizations: A resource dependence perspective*. New York: Harper & Row.

Poppo, L., & Zenger, T. (1998). Testing alternative theories of the firm: Transaction cost, knowledge-based, and measurement explanations for make-or-buy decisions in information services. *Strategic Management Journal*, *19*, 853–877.

Porter, M. E. (1985). *Competitive advantage*. New York: Free Press.

Priem, R., & Butler, J. (2001). Is the resource-based "view" a useful perspective for strategic management research? *Academy of Management Review*, *26*(1), 22–40.

Qian, G., & Li, L. (2003). Profitability of small- and medium-sized enterprises in high-tech industries: The case of the biotechnology industry. *Strategic Management Journal*, *24*(9), 881–887.

Ray, G., Barney, J., & Muhanna, M. (2004). Capabilities, business processes, and competitive advantage: Choosing the dependent variable in empirical tests of the resource-based view. *Strategic Management Journal*, *25*(1), 23–38.

Reger, R., & Huff, A. (1993). Strategic groups: A cognitive perspective. *Strategic Management Journal*, *14*, 103–124.

Robinson, K., & McDougall, P. (1998). The impact of alternative operationalizations of industry structural elements on measures of performance for entrepreneurial manufacturing ventures. *Strategic Management Journal*, *19*, 1079–1100.

Rowe, W., & Morrow, J. (1999). A note on the dimensionality of firm financial performance using accounting, market, and subjective measures. *Canadian Journal of Administrative Sciences*, *16*(10), 58–70.

Rumelt, R., Schendel, D., & Teece, D. (1994). *Fundamental issues in strategy*. Boston, MA: Harvard Business School Press.

Schwab, D. P. (1980). Construct validity in organizational behavior. In: B. Staw & L. Cummings (Eds), *Research in organizational behavior*, (Vol. 2, pp. 3–43). Greenwich CT: JAI Press.

Schwab, D. P. (1999). *Research methods for organizational studies*. Mahwah, NJ: Lawrence Erlbaum.

Shivdasani, A., & Yermack, D. (1999). CEO involvement in the selection of new board members: An empirical analysis. *Journal of Finance*, *54*, 1828–1853.

Shook, C. L., Ketchen, D. J., Cycyota, C. S., & Crockett, D. (2003). Data analytic trends and training in strategic management research. *Strategic Management Journal*, *24*, 1231–1237.

Starbuck, W. (2004). Methodological challenges posed by measures of performance. *Journal of Management and Governance*, *8*(4), 337–343.

Steers, R. (1975). Problems in measurement of organizational effectiveness. *Administrative Science Quarterly*, *20*, 546–558.

Tosi, H., Werner, S., Katz, J., & Gomez-Mejia, L. (2000). How much does performance matter? A meta-analysis of CEO pay studies. *Journal of Management*, *26*(2), 301–339.

Venkatraman, N., & Grant, J. (1986). Construct measurement in organizational strategy research: A critique and proposal. *Academy of Management Review*, *11*(1), 71–87.

Venkatraman, N., & Ramanujam, V. (1986). Measurement of business performance in strategy research: A comparison of approaches. *Academy of Management Review*, *11*, 801–814.

Weinzimmer, L., Nystrom, P., & Freeman, S. (1998). Measuring organizational growth: Issues, consequences, and guidelines. *Journal of Management*, *24*(2), 235–262.

Wiklund, J., & Shepherd, D. (2003). Knowledge-based resources, entrepreneurial orientation, and the performance of small and medium sized businesses. *Strategic Management Journal*, *24*(13), 1307–1314.

Williamson, O. E. (1975). *Markets and hierarchies: Analysis and antitrust implications*. New York: Free Press.

Woo, C., & Willard, G. (1983). Performance representation in business policy research: Discussion and recommendation. Paper presented at Academy of Management meetings, Dallas, TX.

Wright, P., Ferris, S., Hiller, J., & Kroll, M. (1995). Competitiveness through management of diversity: Effects on stock price valuation. *Academy of Management Journal*, *38*, 272–287.