The Cycles of Theory Building in Management Research

Paul R. Carlile
School Of Management
Boston University
Boston, MA 02215
carlile@bu.edu

Clayton M. Christensen
Harvard Business School
Boston, MA 02163
cchristensen@hbs.edu

January 6 2005

Version 6.0

The Cycles of Theory Building in Management Research

Theory thus become instruments, not answers to enigmas, in which we can rest. We don't lie back upon them, we move forward, and, on occasion, make nature over again by their aid. (William James, 1907: 46)

Some scholars of organization and strategy expend significant energy disparaging and defending various research methods. Debates about deductive versus inductive theory-building and the objectivity of information from field observation versus that of large-sample numerical data are dichotomies that surface frequently in our lives and those of our students. Despite this focus, some of the most respected members of our research profession (i.e., Simon (1976), Solow (1985), Hambrick (1994), Staw and Sutton (1995), and Hayes (2002)) have continued to express concerns that the collective efforts of business academics have produced a paucity of theory that is intellectually rigorous, practically useful, and able to stand the tests of time and changing circumstances.

The purpose of this paper is to outline a process of theory building that links questions about data, methods and theory. We hope that this model can provide a common language about the research process that helps scholars of management better understand the roles of different types of data and research, and thereby to build more effectively on each other's work. Our unit of analysis is at two levels: the individual research project and the iterative cycles of theory building in which a researchers attempt to build upon each other's work. The model synthesizes and augments other studies of how communities of scholars cumulatively build valid and reliable theory.[1] It has normative and pedagogical implications for how we conduct research, evaluate the work of others, train our doctoral students, and design our courses.

While many feel comfortable in their understanding of these perspectives, it has been our observation that those who have written about the research process and those who think they understand and practice it proficiently do not yet share even a common language. The same words are applied to very different phenomena and processes, and the same phenomena can be called by many different words. Papers published in reputable journals often violate rudimentary rules for generating cumulatively improving, reliable and valid theory. While recognizing that research progress is hard to achieve at a collective level, we assert here that if scholars and practitioners of management shared and utilized a sound understanding of the process by which theory is built, we could be much more productive in doing research that doesn't just get published, but meets the standards of rigorous scholarship and helps managers know what actions will lead to the results they seek, given the circumstances in which they find themselves.

Our purpose in this paper is *not* to praise or criticize other scholars' work as good theory or bad theory: almost every published piece of research has its unique strengths and shortcomings. We will cite examples of other scholars' research in this paper, but we do so only to illustrate how the theory-building process works. We hope that the model described here might constitute a template and a common language that other scholars might use to reconstruct using how bodies of understanding have accumulated in their own fields.

In the first of the four sections of this paper, we describe a three-step process by which researchers build theory that is at first descriptive, and ultimately normative. Second, we discuss the role that discoveries of anomalies play in the building of better theory, and third, we describe how those who build, evaluate and utilize theories can tell whether they can trust a theory – whether it is valid and applies to the situation in which they find themselves. Finally, we suggest how scholars can engage in *course research* – to design student courses in ways that help faculty researchers build better theory.

I. The Theory Building Process

The building of theory occurs in two major stages–the descriptive stage and the normative stage. Within each of these stages, theory builders proceed through three steps. The the theory-building process iterates through these three steps again and again. In the past, management researchers have quite carelessly applied the term *theory* to research activities that pertain to only one of these steps. Terms such "utility theory" in economics, and "contingency theory" in organization design, for example, actually refer only to an individual step in the theory-building process in their respective fields. We propose that it is more useful to think of the term "theory" as a body of understanding that researchers build cumulatively as they work through each of the three steps in the descriptive and normative stages. In many ways, the term "theory" might better be framed as a verb, as much as it is a noun – because the body of understanding is continuously changing as scholars who follow this process work to improve it.

The Building of Descriptive Theory

The descriptive stage of theory building is a *preliminary* stage because researchers generally must pass through it in order to develop more advanced normative theory. The three steps that researchers who are building descriptive theory utilize are observation, categorization, and association.

Step 1: Observation

In the first step researchers observe phenomena and carefully describe and measure what they see. Careful observation, documentation and measurement of the phenomena in words and numbers is important at this stage because if subsequent researchers cannot agree upon the descriptions of phenomena, then improving theory will prove difficult. Early management studies such as *The Functions of the Executive* (Barnard, 1939) and Harvard Business School cases written in the 1940s and 50s were primarily descriptive work of this genre–and was very valuable. This stage of research is depicted in Figure 1 as the base of a pyramid because it is a necessary foundation for the work that follows. The phenomena being explored in this stage include not just things such as people, organizations and technologies, but process*es* as well. These observations can be done anywhere along the continuum from analysis of huge databases on the one end, to field-based, ethnographic observation on the other.

Without insightful description to subsequently build upon, researchers can find themselves optimizing misleading concepts. As an example: For years, many scholars of inventory policy and supply chain systems used the tools of operations research to derive ever-more-sophisticated optimization algorithms for inventory replenishment. Most were based on an assumption that managers know what their levels of inventory are. Ananth Raman's pathbreaking research of the phenomena, however, obviated much of this work when he showed that most firms' computerized inventory records were broadly inaccurate – even when they used state-of-the-art automated tracking systems (Raman 199X). He and his colleagues have carefully described how inventory replenishment systems work, and what variables affect the accuracy of those processes. Having laid this foundation, supply chain scholars have now begun to build a body of theories and policies that reflect the real and different situations that managers and companies face.

Researchers in this step often develop what we term *constructs.* Constructs are abstractions that help us rise above the messy detail to understand the essence of what the phenomena are and how they operate. Joseph Bower's *Managing the Resource Allocation Process* (1970) is an example of this. His constructs of *impetus* and *context*, explaining how momentum builds behind certain investment proposals and fails to coalesce behind others, have helped a generation of policy and strategy researchers understand how strategic investment decisions get made. Economists' concepts of "utility" and "transactions costs" are constructs – abstractions developed to help us understand a class of phenomena they have observed. We would not label the constructs of utility and transactions cost as theories, however. They are *part* of theories – building blocks upon which bodies of understanding about consumer behavior and organizational interaction have been built.

Step 2: Classification

With the phenomena observed and described, researchers in the second stage then classify the phenomena into categories. In the descriptive stage of theory building, the classification schemes that scholars propose typically are defined by the attributes of the phenomena. Diversified vs. focused firms, and vertically integrated vs. specialist firms are categorization examples from the study of strategy. Publicly traded vs. privately held companies is a categorization scheme often used in research on financial performance. Such categorization schemes attempt to simplify and organize the world in ways that highlight possibly consequential relationships between the phenomena and the outcomes of interest.

Management researchers often refer to these descriptive categorization schemes as *frameworks* or *typologies*. Burgelman & Sayles (1986), for example, built upon Bower's (1970) construct of context by identifying two different types of context–organizational and strategic.

Step 3: Defining Relationships

In the third step, researchers explore the association between the category-defining attributes and the outcomes observed. In the stage of descriptive theory building, researchers recognize and make explicit what differences in attributes, and differences in the magnitude of those attributes, *correlate* most strongly with the patterns in the outcomes of interest. Techniques such as regression analysis typically are useful in defining these correlations. Often we refer to the output of studies at this step as *models*.

Descriptive theory that quantifies the degree of correlation between the category-defining attributes of the phenomena and the outcomes of interest are generally able to make probabilistic statements of association representing average tendencies. For example, Hutton, Miller and

Skinner (2000) have examined how stock prices respond to earnings announcements. They coded types of words and phrases in the statements as explanatory variables in a regression equation, with the ensuing change in equity price as the dependent variable. This analysis enabled the researchers then to assert that, on average across the entire sample of companies and announcements, delivering earnings announcements in a particular way would lead to the most favorable (or least unfavorable) reaction in stock price. Research such as this is important descriptive theory. However, at this point it can only assert *on average* what attributes are associated with the best results. A specific manager of a specific company cannot yet know whether following that average formula will lead to the hoped-for outcome in her specific situation. The ability to know what actions will lead to desired results for a specific company in a specific situation awaits the development of normative theory in this field, as we will show below.

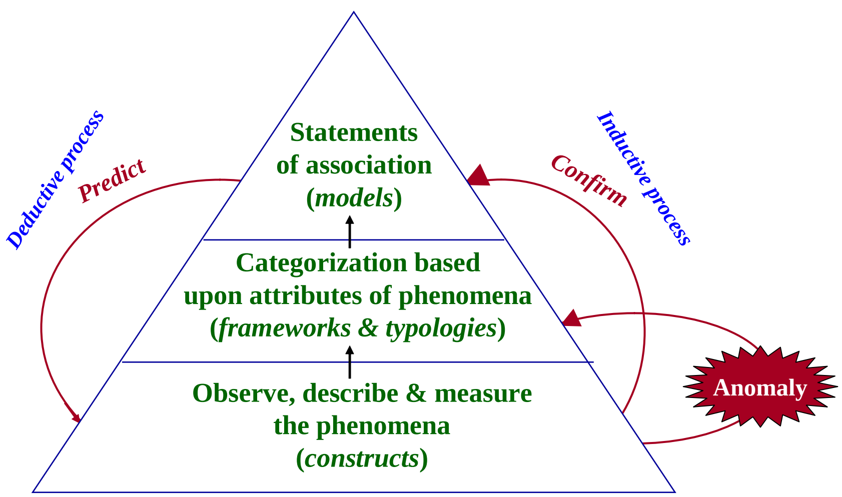
How Theory is Improved within the Descriptive Stage

When researchers move from the bottom to the top of the pyramid in these three steps–observation, categorization and association, and in so doing give us constructs, frameworks and models–they have followed the *inductive* portion of the theory building process. Researchers can then get busy improving these theories by cycling from the top down to the bottom of this pyramid in the *deductive* portion of the cycle – seeking to "test" the hypotheses that had been inductively formulated. This most often is done by exploring whether the same correlations exist between attributes and outcomes in a different set of data than the data from which the hypothesized relationships were induced. When scholars test a theory on a new data set (whether the data are numbers in a computer, or are field observations taken in a new context), they sometimes find that the attributes of the phenomena in the new data do indeed correlate with the outcomes as predicted. When this happens, this "test" confirms that the theory is of use under the conditions or circumstances observed.[2] However, researchers who stop at this point simply return the model to its place atop the pyramid *tested but unimproved.*

It is only when an anomaly is identified–an outcome for which the theory can't account–that an opportunity to improve theory occurs. As Figure 1 suggests, discovery of an anomaly gives researchers the opportunity to revisit the foundation layers in the theory pyramid–to define and measure the phenomena more precisely and less ambiguously, or to cut it into alternative categories–so that the anomaly *and* the prior associations of attributes and outcomes can all be explained. In the study of how technological innovation affects the fortunes of leading firms, for example, an early attribute-based categorization scheme was radical vs. incremental innovation. The statements of association that were built upon it concluded that the leading established firms on average do well when faced with incremental innovation, but they stumble in the face of radical change. But there were anomalies to this generalization – established firms that successfully implemented radical technology change. To account for these anomalies, Tushman & Anderson (1986) offered a different categorization scheme, competency-enhancing vs. competency-destroying technological changes. This scheme resolved many of the anomalies to the prior scheme, but subsequent researchers uncovered new ones for which the Tushman-Anderson scheme could not account. Henderson & Clark's (1990) categories of modular vs. architectural innovations; Christensen's (1997) categories of sustaining vs. disruptive technologies; and Gilbert's (2001) threat-vs.-opportunity framing each uncovered and resolved anomalies for which the work of prior scholars could not account. This body of understanding has improved and become remarkably useful to practitioners and subsequent scholars (Adner, 2003; Daneels, 2005) *because* these scholars followed the process in a disciplined way. They articulated theories that could be falsified – that could yield anomalies. Subsequent scholars then uncovered what these anomalies were, and resolved them by slicing the phenomena in different ways and articulating new associations between the category-defining attributes and the outcome of interest.

Figure 1

The Process of Building Theory



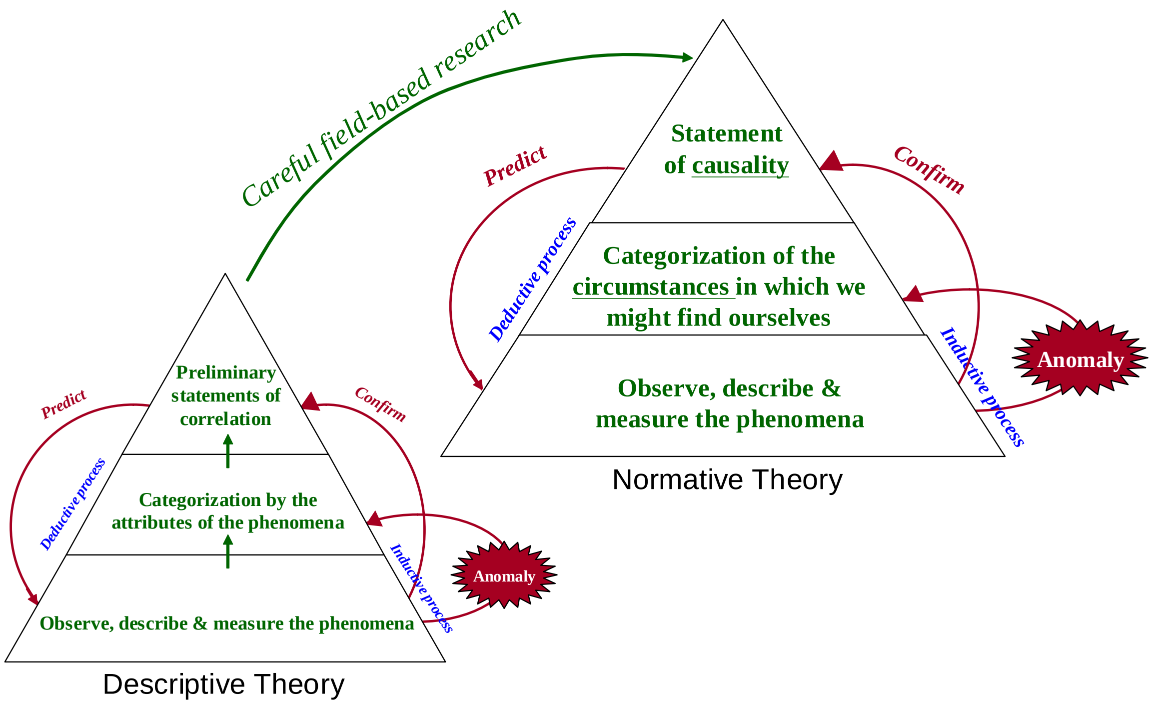
In contrast to many debates about the virtues of deductive and inductive methods, this suggests that these are two sides to the same pyramid. Every complete lap around the theory-building pyramid consists of an inductive side and a deductive side. Theory building efforts stall when researchers drop the baton and declare victory having run only half of a lap around the theory pyramid.[3]

Descriptive theory-building efforts typically categorize by the attributes of the phenomena because attributes are easiest to observe and measure. Likewise, correlations between attributes and outcomes are easiest to hypothesize and quantify through techniques such as regression analysis. Kuhn (1962) observed that confusion and contradiction typically are the norm during descriptive theory-building. This phase is often characterized by a plethora of categorization schemes because the phenomena generally have many different attributes. The sequence of studies of technology change cited above is an illustration of such a plethora. Often, in this phase, no model is irrefutably superior: Each seems able to explain anomalies to other models, but suffers from anomalies to its own.

The Transition from Descriptive to Normative Theory

The confusion and contradiction that often accompany descriptive theory become resolved when careful researchers – often through detailed empirical and ethnographic observation – move beyond statements of correlation to define what *causes* the outcome of interest. As depicted in Figure 2, they leap across to the top of the pyramid of normative theory, whose capstone is a statement of what *causes* the outcome of interest, not just what is correlated with it. Their understanding of causality enables researchers to assert what actions managers ought to take, in order to get the results they need. For reasons noted below, normative theory has much greater predictive power than descriptive theory does.[4]

Figure 2:
The Transition from Descriptive Theory to Normative Theory



Normative theory, like its descriptive predecessor, still needs to be improved – and researchers do this by following the same steps that were used in the descriptive stage. Hypothesizing that their statement of causality is correct, they cycle deductively to the bottom of the pyramid to test the causal statement: If we observe these actions being taken, these should be the outcomes that we observe. When they encounter an anomaly, they then delve back into the lower levels of the pyramid. Sometimes they can resolve anomalies by developing more accurate, less ambiguous ways to define and measure the phenomena. Often they account for the anomalies by revisiting the categorization stage. Rather than using schemes based on attributes of the phenomena, however, in building normative theory researchers categorize *the different situations or circumstances* in which managers might find themselves. They do this by asking, when they encounter an anomaly, "What was it about the situation in which those managers found themselves, that caused the causal mechanism to yield a different result? By asking this question as they cycle up and down the pyramid of normative theory, anomaly-seeking researchers will ultimately define a relatively complete set of the situations or circumstances in which managers might find themselves when pursuing the outcomes of interest.[5] This allows researchers to make contingent statements of causality – to show how and why the casual mechanism results in a different outcome, in the different situations. A normative theory that is built upon well-researched categories of circumstances can help a manager predict accurately what actions will and will not lead to the desired result, given the circumstance in which she finds herself. The relatively accurate, circumstance-contingent predictability of normative theory enables managers to know, in other words, what they *ought* to do.[6]

The history of research into manned flight is a good way to visualize how this transition from descriptive to normative theory occurs, and how it is valuable. During the middle ages, would-be aviators did their equivalent of best-practices research and statistical analysis. They observed the many animals that could fly well, and compared them with animals that could not. The vast majority of the successful fliers had wings with feathers on them; and almost all of those that couldn't fly had neither of these attributes. This was quintessential descriptive theory. Pesky outliers like ostriches had feathered wings but couldn't fly; bats had wings without feathers and were very good at it; and flying squirrels had neither and got by. But the R^2 was so high that aviators of the time copied the seemingly salient characteristics of the successful fliers in the belief that if they copied the characteristics of the "best practices" fliers, they could fly, too. So they fabricated wings, glued feathers on them, jumped off cathedral spires, and flapped hard. It never worked. For centuries they sought to fly by trying harder – assuming that the prior aviators had failed because they had bad wing designs; hadn't bulked up their muscles enough; or hadn't flapped hard enough. There were substantial disagreements about the categorization scheme, too – which of the birds' attributes truly enabled flight, and which didn't. For example, Roger Bacon wrote an influential paper asserting that the differentiating characteristic was birds' hollow bones (Clegg, 2003). Because man had solid bones, Bacon reasoned, we could never fly. He then proposed several designs of machines that could flap their wings with sufficient power to overcome the disadvantage of solid human bones. But it still never worked. Armed only with the correlative statements of descriptive theory, aviators kept killing themselves.

Then through his careful study of fluid mechanics Daniel Bernoulli identified a shape that we call an airfoil – a shape that, when it cuts through air, creates a mechanism that we call lift. Understanding this causal mechanism, Bernoulli's Principle, made flight possible. But it was not yet predictable. In the language of this paper, the theory predicted that aviators would fly successfully when they built machines with airfoils to harness lift. But while they sometimes flew successfully, occasionally they did not. Crashes were anomalies that Bernoulli's theory could not explain. Discovery of these anomalies, however, allowed the researchers to revisit the categorization scheme. But this time, instead of slicing up the world by the attributes of the good and bad fliers, researchers categorized the world by circumstance – asking the question, "What was it about the circumstance that the aviator found himself in that caused the crash?" This then enabled them to improve equipment and techniques and articulate circumstance-contingent statements of causality: "This is how you should normally fly the plane. But when you get in this situation, you need to fly it differently in order to get the desired outcome. And when you get in this other situation, don't even try to fly. It is impossible."

When their careful studies of anomalies allowed researchers to identify the set of circumstances in which aviators might find themselves, and then modified the equipment or developed piloting techniques that were appropriate to each circumstance, manned flight became not only possible, but much more predictable. This is how this body of understanding about human flight transitioned from descriptive to normative theory. It was the discovery of the fundamental *causal* mechanism that made flight possible. And it was the categorization of the salient circumstances that enabled flight to be made more predictable.

The world of managers is unlikely ever to become perfectly predictable, of course. Managers, like pilots, will likely continue to find themselves in never-before-encountered situations for which adequate rules and equipment have not yet been created. Complicated human factors and group dynamics also militate against perfect predictability. But even here, the work of scholars of group dynamics such as Richard Hackman (198x) has done much to help us understand what behaviors lead to what results and why – and how the result might differ by circumstance.[7]

On the Importance of the Categorization Step

Several prominent scholars have examined the improvement in predictability that accompanies the transition from the attribute-based categorization of descriptive theory, to the circumstance-based categorization of normative theory. Consider, for example, the term "Contingency Theory" – a concept born of Lawrence & Lorsch's (1967) seminal work. They showed that the best way to organize a company depended upon the circumstances in which the company was operating. In our language, contingency is not a theory *per se*. Rather, contingency is the categorization scheme, and is a crucial element of *every* normative theory. Rarely do we find one-size-fits-all answers to every company's problem.

Glaser and Strauss's (1967) treatise on "grounded theory" actually is a book about categorization. Their term *substantive theory* corresponds to the attribute-bounded categories of *descriptive theory*. And their concept of *formal* theory matches our definition of normative theory that employs categories of circumstance.

Management fads often are created when a researcher studies a few successful companies, finds that they share certain characteristics, concludes that he has seen enough, and then skips the categorization step entirely by writing a book asserting that if all managers would imbue their companies with the characteristics of these successful companies, they would be similarly successful. When managers then apply the formula, some generally find that it doesn't work in their companies. This casts a pall on the idea. Some faddish theories aren't uniformly bad. It's just that their authors were so eager for their theory to apply to everyone that they never took the care to to figure out the circumstances in which their statement of causality would lead to success, and when it would not. Efforts to identify "the best practices of successful companies" almost uniformly suffer from this problem.[8]

Unfortunately, it is not just authors-for-profit of management books that contribute to the problem of publishing theory whose applicability is not guided by categorization. Reading a typical scholarly management journal today can be depressing – because the vast majority of published papers devote few of their column inches to categorization. When the existence of different categories is noted, often they are "handled" with dummy variables or by omitting the outliers – as if maximizing R^2, rather than getting the categories clearly characterized, is the hallmark of a good theory.

Other well-intentioned academics unwittingly contribute to the problem by articulating tight "boundary conditions" outside of which they claim nothing. Delimiting the applicability of a theory to the specific time, place, industry and/or companies from which the conclusions were drawn in the first place is a mutation of one of the cardinal sins of research – sampling on the dependent variable. In order to be useful to managers and to future scholars, researchers need to help managers understand the circumstance that they are in. Almost always, this requires that they also be told about the circumstances that they are *not* in. In other words, the process of getting the categories right is an ever-challenging but always important step in theory building.

Some management researchers are convinced that human-laden social systems are so multi-faceted and dynamic that meaningful simplification into a few categories of circumstance is impossible. They assert that managers' world is so complex that there are an infinite number of situations in which they might find themselves. Indeed, this seemingly infinite complexity of categories very nearly characterizes the descriptive theory phase in some fields. But normative theory generally is not so confusing. Researchers in the normative theory phase resolve confusion by abstracting up from the detail to define a few categories – typically two to four – that comprise *salient* circumstances.

As noted above, Thomas Kuhn (1962) discussed in detail the transition from descriptive to normative theory in his study of the emergence of scientific paradigms. He described a preliminary period of confusion and debate in theory building, which is an era of descriptive theory. His description of the emergence of a paradigm corresponds to the transition to normative theory described above. Even when a normative theory achieves the status of a broadly believed paradigm, it continues to be improved through the process of discovering anomalies, as we describe above.[9] Indeed, the emergence of new phenomena – which probably happens more frequently in competitive, organizational and social systems than in the natural sciences – ensures anomaly-seeking researchers of management that there probably will always be additional productive laps up and down the theory pyramid that they can run.

Finding the Salient Boundaries of Circumstance-based Categories

If circumstance-defined categorization is so critical to the building of normative theory, how do researchers decide what boundaries best define the categories, and what potential definitions of boundaries are not salient to accurate prediction and understanding? Returning to our account of aviation research, the boundaries that defined the salient categories of circumstance are determined by the necessity to pilot the plane differently. If a different circumstance does not require different methods of piloting, then it is not a meaningful category. We propose that the same principle defines the salience of category boundaries in management theory. If managers find themselves in a circumstance where they must change actions or organization in order to achieve the outcome of interest, then they have crossed a salient boundary between categories.

The Value of Anomalies in Building Better Theory

As indicated before, when researchers in both the descriptive and normative stages cycle down from the top of the pyramid using statements of association or causality to predict what they will see at the foundation, they often observe something that the theory did *not* lead them to expect; thus identifying an *anomaly*—something the theory could not explain. Such discoveries force theory builders to cycle back into the description / measurement and categorization stages with puzzles such as "there's something else going on here" or "these two things that we thought were different, really aren't." The results of this effort typically can include: 1) more accurately describing and measuring what the phenomena are and are not; 2) changing the definitions by which the phenomena or the circumstances are categorized – adding or eliminating categories or defining them in different ways; and/or 3) articulating a new theoretical statement of what is associated with, or causes what, and why, and under what circumstances. The objective of this process is to revise theory so that it still accounts for both the anomalies identified and the phenomena that were previously well-explained.

Anomalies are valuable in theory building because *the discovery of an anomaly is the enabling step to less ambiguous description and measurement, and to identifying and improving the categorization scheme in a body of theory* – which are the keys to being able to apply the theory with predictable results. Researchers whose goal is to "prove" a theory's validity often view discovery of an anomaly as failure – and therefore search for reasons to exclude outlying data points in order to get more significant measures of statistical fit. There typically is more information in the points of outlying data than in the ones that fit the model well, however, because understanding the outliers or anomalies is generally the key to discovering problems in definition & measurement, and in formulating better categorization schemes. Journal editors and peer reviewers whose objective is to improve theory should therefore embrace papers that seek to surface and resolve anomalies, and view with less interest studies that seek to avoid them.

Indeed, productive theory-building research is almost invariably prompted or instigated by an anomaly or a paradox (Poole & Van de Ven, 1989). The research that led to Michael Porter's (1991) *Competitive Advantage of Nations* is an example. Before Porter's work, the theory of international trade was built around the notion of comparative advantage. Nations with inexpensive electric power, for example, would have a competitive advantage in those products in which the cost of energy was high; those with low labor costs would enjoy an advantage in making and selling products with high labor content; and so on. Porter saw anomalies for which this theory could not account. Japan, with little iron ore and coal, became a successful steel producer. Italy became the world's dominant producer of ceramic tile even though it had high electricity costs and had to import much of the clay used in making the tile. Porter resolved these anomalies by categorizing the world into two circumstances – situations in which a factor-based advantage exists, and those in which it does not. In the first situation the reigning theory of comparative advantage still has predictive power. But in the latter circumstance, Porter's theory of competitive industrial clusters explained the phenomena that had been anomalous to the prior theory. Porter's theory is normative because it gives planners clear guidance about what they should do, given the circumstance in which they find themselves. The government of Singapore, for example, attributes much of that country's prosperity to the guidance of Porter's theory.[10]

Yin (1984) distinguishes between *literal* replications of a theory, versus *theoretical* replications. A literal replication occurs when the predicted outcome is observed. A theoretical replication occurs when an unusual outcome occurs, but for reasons that can be explained by the model. This means that we must dive much more deeply before we label an "exception" to a theory's predictions as an anomaly. For example, the observation that airplanes fly is an exception to the general assertion that the earth's mass draws things down toward its core. Does this exception disprove the theory of gravity? Of course not. While falling apples and flaming meteors are literal replications of the theory, manned flight is a theoretical replication. It is a different outcome than we normally would expect, but Bernoulli's Principle explains why it can occur. An anomaly is an outcome that is neither a literal or theoretical replication of a theory.

How to Design Anomaly-Seeking Research

Although some productive anomalies might be obvious from the outset, often the task of theory-building scholars is to design their research to *maximize the probability that they will be able to identify anomalies*. The purpose of the following paragraphs is to help researchers to be more productive in theory-building by defining research questions that focus on anomalies, and then by creating research designs that maximize the probability of discovering important anomalies as they execute the research. We conclude this section by describing how literature reviews might be structured to help readers understand how knowledge has accumulated in the past, and thereby to position a paper more compellingly in its relevant stream of scholarship.

1. Anomaly-Seeking Research Questions

Anomaly-seeking research enables new generations of researchers to pick up even well-accepted theories, and to run the theory-building cycle again – adding value even to research that already has earned broad praise and acceptance. Consider Professor Porter's (1991) research mentioned above. In Akron, Ohio there was a powerful cluster of tire manufacturers whose etiologies and interactions could be explained well by Porter's theory. That cluster subsequently vaporized – in part because of the actions of a company, Michelin, that operated outside of this cluster (Sull, 2000). This anomaly suggests that there must situations in time or space in which competing within a cluster is competitively important. Certain subsequent developments might render clusters less important. When an improved categorization scheme for this theory emerges from the work of Sull and others, the community of strategy scholars and policy makers will have an even clearer sense for when the competitive crucible of clusters is critical for developing capabilities, when it is not, and why.

In this spirit, we outline below some examples of potentially productive questions that future researchers could pursue, and potentially challenge the categorization schemes used in important bodies of management research:

- When might process re-engineering or lean manufacturing be *bad* ideas?
- When could sourcing from a partner or supplier something that is not your core competence lead to *disaster*?
- Are there circumstances in which pencil-on-paper methods of vendor management yield better results than using supply-chain management software?
- In retailing, when and why is a one-stop-shopping or "portal" strategy effective and when would we expect firms using focused specialist strategies to gain the upper hand?

- When are time-based competition and mass customization likely to be critical and when might they be competitively meaningless?

- Are SIC codes the right categories for defining "relatedness" in diversification research?

- When should acquiring companies integrate a firm they have just purchased into the parent organization, and when should they keep it separate?

- What is the circumstance in which Chandler's (1977) "visible hand" of managerial capitalism is critical to an enterprise's success, and in what circumstance is Adam Smith's (1776) "invisible hand" the best coordinating mechanism for enterprise-building?

Anomaly-seeking research *always* is focused on the categorization step in the pyramid. Unfortunately, the category boundaries in much research (such as SIC codes) seem to be defined by the availability of data, rather than their salience to the underlying phenomena or their relation to the outcome. Questioning the sufficiency of such categories is almost always a productive path for building better theory. "When doesn't this work?" and "Under what conditions might this gospel be bad news?" are simple questions that can yield breakthrough insights – and yet too few researchers have the instinct to ask them.

2. The Lenses of Other Disciplines

One of Kuhn's (1962) most memorable observations was that the anomalies that led to the toppling of a reigning theory or paradigm almost invariably were observed by researchers whose backgrounds were in different disciplines than those in which the leaders in the field traditionally had been trained. Their beliefs about what was and was not possible seemed to shape so powerfully what the adherents to the prior theory could and could not see that they often went to their graves denying the existence or relevance of the anomalous phenomena that led to the creation of improved theory. Researchers from different disciplines generally use different methods and have different interests toward their object of study. Such differences often allow them to see things that might not be recognized or might appear inconsequential to an insider.

It is not surprising, therefore, that many of the most important pieces of breakthrough research in the study of management, organization and markets have come from scholars who stood astride two or more academic disciplines. Porter's (1980, 1985, 1991) work in strategy, for example, resulted from his having combined insights from business policy and industrial organization economics. The insights that Robert Hayes and his colleagues (1980, 1984, 1985, 1988) derived about operations management combined insights from process research, strategy, cost accounting and organizational behavior. Baldwin & Clark's (2000) insights about modularity were born at the intersection of options theory in finance with studies of product development. Clark Gilbert (2001) looked at Christensen's (1997) theory of disruptive innovation through the lenses of prospect theory and risk framing (Kahnemann & Tversky 1979, 1984), and saw explanations of what had seemed to be anomalous behavior, for which Christensen's model could not account.

3. Studying the Phenomena within the Phenomena

The third method to increase the probability that researchers will identify anomalies is to execute *nested* research designs that examine different levels of phenomena. Rather than study just industries *or* companies *or* divisions *or* groups *or* individuals, a nested research design entails studying how individuals act and interact within groups; and how the interaction amongst groups and the companies in which they are embedded affect the actions of individuals. Many anomalies will only surface while studying second-order interactions across levels within a nested design.

The research reported in Johnson & Kaplan's *Relevance Lost* (1987) is a remarkable example of the insights gained through nested research designs. Most prior researchers in managerial accounting and control had conducted their research at a single level—the numbers printed in companies' financial statements. Johnson and Kaplan saw that nested beneath each of those printed numbers was a labyrinth of political, negotiated, judgmental processes that could systematically yield inaccurate numbers. This led to their devising methods of activity-based costing, which can better assure that numbers in financial statements more accurately represent actual costs.

Spear and Bowen (1999) developed their path-breaking insights of the Toyota Production System through a nested research design. Researchers in the theory's descriptive stage had studied Toyota's production system at single levels. They documented visible artifacts such as minimal inventories, *kanban* scheduling cards and rapid tool changeovers. After comparing the performance of factories that did and did not possess these attributes, early researchers asserted that if other companies would use these same tools, they could achieve similar results (see, for example, Womack *et.al.*, 1990). The anomaly that gripped Spear and Bowen was that when other firms used these artifacts, they still weren't able to achieve Toyota's levels of efficiency and improvement. By crawling inside to study how individuals interacted with individuals, in the context of groups interacting with other groups, within and across plants within the company and across companies, Spear and Bowen were able to go beyond the correlative statements of descriptive theory, to articulate the fundamental causal mechanism behind the Toyota system's self-improving processes.[11]

4. Observing and Comparing a Broad Range of Phenomena

The fourth mechanism for maximizing the probability of surfacing an anomaly is to examine, in the deductive half of the cycle, a broader range of phenomena than prior scholars have done. The broader the range of outcomes, attributes and circumstances that scholars examine at the base of the pyramid, the higher the probability that they will identify the salient boundaries among the categories. As an example, Chesbrough's (1999) examination of Japanese disk drive makers (which Christensen had not focused upon in his study) enabled Chesbrough to surface anomalies for which Christensen's theory of disruptive technology could not account—leading to an even better theory that explains how horizontally and vertically integrated companies can manage disruption differently than focused companies can.

Anomaly-Seeking Research and the Cumulative Structure of Knowledge

When we ask new faculty candidates who as doctoral students have been trained in methods of modeling and data analysis to position their work within a stream of scholarship, we observe that they readily recite long lists of articles in "the literature." But they struggle when asked which scholar's work resolves anomalies to prior scholars' theories; whose results contradicted whose, and why. Many lists of prior publications are simply lists, sometimes lifted from prior authors' lists of prior articles. Even though their professor-advisers purport to understand how theory is built, few of their doctoral students have been taught to organize citations in a way that describes the laps that prior researchers have taken, and they cannot give readers a sense for how theory has or has not been cumulatively built to date. Rather, after doffing the obligatory cap to prior research, most doctoral students get busy testing their hypotheses in the belief that if nobody has tested these particular ones before, using novel analytical methods on a new data set, they must be breaking new ground. Then as faculty members they institutionalize these research habits when they train new doctoral students and review colleagues' papers for publication.

We suggest that in the selection of research questions and the design of research methods, researchers might map the literature on a large sheet of paper in the format of Figure 2 above, following the method of Gilbert (2005). They then should answer questions like these:

- Is this body of theory in the descriptive or normative stage?

- What anomalies have surfaced in prior authors' work, and which pieces of research built on those by resolving the anomaly? What are the generations of categorization schemes that prior scholars have proposed, and how have these schemes improved through this process?

- At what step must I be positioning my work in order to move this collective enterprise forward? Do I need to be at the base of the pyramid defining constructs that abstract from the detail of the phenomena, so that students of this body of knowledge can better understand what really is going on? Is the greatest need to strengthen this foundation by offering ways to examine and measure the phenomena more unambiguously? Am I resolving an anomaly by suggesting that prior scholars haven't categorized things correctly? Am I running half a lap or a complete cycle, and why?

Researchers who attempt this often will be shocked to see how difficult it is to map how knowledge has accumulated within a given sub-field. In many cases, it simply hasn't summed up to much, as the critics cited in our first paragraph have observed.

We suggest that the three-step pyramids of descriptive and normative theory building might constitute a generic map that could help scholars organize the collective efforts of researchers within their fields. The curriculum of doctoral seminars might be organized in this manner, so that students are brought through the past into the present in ways that help them visualize the next steps required to build better theory. Literature reviews, if constructed in this way at the beginning of papers, would help readers position the work in the context of this stream, in a way that adds much more value than listing articles in alphabetical or chronological order.

Theory-building vs. Theorizing

A significant portion of researchers in certain fields of management who focus their efforts on the building of mathematical models start their work at the top of the theory pyramid and stay there, because their optimization models are not designed to yield the sort of empirical anomalies that we have discussed above. These bodies of work tend to improve as ever-more clever mathematicians devise ways to include more variables in their models, or discover how to derive results while making more "realistic" simplifying assumptions than prior scholars had been able to do. Because these improvements tend to be achieved through conceptual rather than empirical falsification, we term research of this genre as *theorizing* rather than *theory-building*. In the spirit of this paper's attempt to lay a solid descriptive foundation for a theory of theory building, we feel it is important to assign a different term to the process that pure modelers employ. Theorizing is a fundamentally different enterprise than is the theory-building process described here, in which the search for empirical anomalies plays such an intrinsic role.

Establishing the Validity and Reliability of Theory

A primary concern of every consumer of management theory is to understand where it applies, and where it does not apply. Yin (1984) helps us with these concerns by defining two types of validity for a theory – internal and external validity – that help us guage whether and when we can trust it. In this section we'll discuss how these concepts relate to our model of theory building, and describe how researchers can make their theories more valid on both of these dimensions.

Internal Validity

A theory's *internal* validity is the extent to which: 1) its conclusions are unambiguously drawn from its premises; and 2) the researchers have ruled out all plausible alternative explanations that might link the phenomena with the outcomes of interest. The best way we know to establish the internal validity of a theory is to examine the phenomena through the lenses of as many disciplines and parts of the company as possible – because the plausible alternative explanations almost always are found by examining the workings of another part of the company, as seen through the lenses and tools of other academic disciplines.

An illustration of the value that multi-lens research brings to theory can be seen in accounts of the about-face that Intel engineered in the early 1980s, as it exited the Dynamic Random Access Memories (DRAM) business and threw all of its resources behind its microprocessor strategy. Some accounts of this impressive achievement attribute its success to the leadership and actions of its visionary leaders, Gordon Moore and Andy Grove (see, for example, Yoffie *et.al.* 2002). Burgelman's (2002) careful reconstruction of the role of each functional group in those years of transition, however, reveals a very different explanation of how and why Intel was able to make this transition. He showed how methods of product costing, sales forecasting, incremental problem-solving and saleforce compensation interacted within Intel's resource allocation process to execute this change on auto-pilot, with little regard to the decisions of the senior-most management. There was a *correlation* between the presence of powerful, visionary leaders and this major strategic change, but Burgelman's multi-lens approach enabled him to show that it wasn't the cause.

The relational perspective of Black, Repenning & Carlile (2005) similarly enhanced the internal validity of Barley's famous (1986) CT scanning study, giving us a more comprehensive understanding of what causes what and why. Barley originally had shown that four different circumstances resulted in the implementation of the same technology in four settings – and concluded that context drives everything. By formalizing the relations (not just the attributes) among the actors involved and with the new CT scanning technology they used, Black *et.al.* revealed the relational properties among actors and the technology that explain all four of the circumstances Barley had observed. Understanding the critical relational properties amongst the people and technologies can give managers not just a more comprehensive sense of causality, but a roadmap for how to implement strategies that predictably lead to the needed results.

When there's a possibility that another researcher could say, "Wait a minute. There's a totally different explanation for why this happened," we cannot be assured of a theory's internal validity. Scholars who examine the phenomena and outcomes of interest through the lenses of all potentially relevant perspectives, can either incorporate what they learn into their explanations of causality, or rule out other explanations so that theirs is the only plausible one left standing.

External Validity

The *external* validity of a theory is the extent to which a relationship that was observed between phenomena and outcomes in one context can be trusted to apply in different contexts as well. Many researchers have come to believe that a theory's external validity is established by 'testing' it on different data sets. This can *never* conclusively establish external validity, however – for two reasons. First, researchers cannot test a theory on every conceivable data set. Second, data only exists about the past. How can we be sure a model applies in the present or future, before there is data to test it on?

To illustrate this problem, consider Christensen's experience after publishing the theory of disruptive innovation. This normative theory had been inductively derived through careful empirical analyses of the history of the disk drive industry. It asserted that there are two circumstances – sustaining and disruptive situations – in which innovating managers might find themselves. Then it defined a causal mechanism – the functioning of the resource allocation process in response to the demands of customers and financial markets – that caused leading incumbent firms and entrants to succeed or fail at different types of innovations in those circumstances.

Those who read Christensen's early papers instinctively wondered, "Does this theory apply outside the disk drive industry?" To address these concerns when writing *The Innovator's Dilemma*, Christensen (1997) sought to establish the generalizability or external validity of the theory by "testing" it on data from as disparate a set of industries as possible – including hydraulic excavators, department stores, steel, computers, motorcycles, diabetes care, accounting software, motor controls and electric vehicles. Despite the variety of industries in which the theory seemed to have explanatory power, executives from industries that weren't specifically studied kept asking, "Does it apply to health care? Education? Financial services?" When Christensen published additional papers that applied the model to these industries, the response was, "Yes, I see that. But does it apply to telecommunications? Database software? The German economy?" The killer question, from an engineer in the disk drive industry, was, "It clearly applies to the *history* of the disk drive industry. But does it apply to its *future* as well? Things are very different now." As these queries illustrate, it is simply impossible to establish the external validity of a theory by testing it on data. There will always be another set upon which it hasn't yet been tested, and the future will always lie just beyond the reach of data.

External validity can only be established through categorization. We can say that a normative theory is externally valid when the categories of circumstance are mutually exclusive and collectively exhaustive. Mutually exclusive categorization allows managers to say, "I am in this circumstance and not any of those others." And collectively exhaustive categorization would assure us that all situations that managers might find themselves in with respect to the phenomena and outcomes of interest, are accounted for in the theory. No theory's categorization scheme is likely to achieve permanent status of mutually exclusive and collectively exhaustive, of course. But the refinements that come from cycles of anomaly-seeking research can asymptotically improve theory towards that goal.

Sample Size and Validity

Methods of measuring statistical significance certainly show that the larger the sample size, the more certain we can be of a model's *internal* validity. Some are also tempted to think that a theory derived from a large data set representing an entire population of companies would have greater *external* validity than a theory derived from case studies of a limited number of situations within that population. But this is not the case. When the unit of analysis is a population of companies, the researcher can be specific only about the entire population of companies. Some managers will find that following the formula that works best on average for the population, works best in their situation as well. However, sometimes the course of action that is optimal on average will *not* yield the best outcome in a specific situation. Hence, researchers who derive a theory from statistics about a population still need to establish external validity through circumstance-based categorization.[12]

What types of data can be trusted to yield reliable theory?

In addition to internal and external validity, a third measure of a theory's quality is its reliability – the extent to which another researcher could derive the same conclusions from the same observations. The dichotomy between subjectivity and objectivity is often used as a cleavage point to judge the reliability of data and theories derived therefrom – with many seeing numerical, objective data as more trustworthy than subjective data. But where does quantitative, "objective" data come from? The data used in many research projects comes from companies' financial statements, for example. Is this objective? Johnson & Kaplan (1987) showed quite convincingly that the numbers representing revenues, costs and profits in financial statements are the result of processes of estimation, negotiation, debate and politics that can produce grossly inaccurate reflections of true cost and profit. Even the "hardest" of numbers, such as those measuring prices and product performance, really are after-the-fact proxy manifestations of the prioritizations, fudging measurements, exaggerations and negotiations that occurred before a number appeared to represent all of those things.

The healthiest and most accurate mindset for researchers is that nearly *all* data – whether presented in the form of large data sample analysis on one extreme, or an ethnographic description of behavior on the other – are subjective. Numerical and verbal data alike are abstractions from a much more complex reality, out of which the researcher attempts to pull the most salient variables or patterns for examination. Whereas the subjectivity of data from field-based, ethnographic research is glaringly apparent, the sources of subjectivity are hidden behind numerical data and cannot be reviewed.[13] There should be no smugness amongst quantitative researchers about the alleged objectivity of their data, and no defensiveness amongst field researchers about the subjectivity of theirs. We are all in the same subjective boat, and are obligated to do our best to be humble and honest with ourselves and our colleagues about where our data comes from as we participate individually within and collectively across the theory building cycle.

Much like theory, the only way we can judge the value of data is by their usefulness in helping us understand how the world works, identifying categories, making predictions and surfacing anomalies.

Implications for Course Design

Schools of management generally employ two methods of classroom instruction: case-based and lecture-based classes. These are descriptive categorizations of the phenomena. Attempts to assess which method of instruction is associated with the best outcomes are fraught with anomaly. We suggest that there is a different, circumstance-based categorization scheme that may constitute a better foundation of a theory of course design: Whether the instructor is using the course to *develop* theory, or to help students practice the *use* of theory.

When designing a course on a subject about which normative theory has not yet emerged, designing the course to move up the *inductive* side of the theory pyramid can be very productive. For example, in 1998 Harvard Business School professor Kent Bowen decided to create a course on owning and running small businesses, because many MBA graduates end up doing just that. Then discovering that little had been written about how to run low-tech, slow-growth companies, he tackled the problem with an inductive course design. He first wrote a series of cases that simply described what managers in these sorts of companies worry about and do. The purpose of each case discussion was to help the professor and students to understand the phenomena thoroughly. After a few classes, Bowen paused, and orchestrated a class discussion to define patterns in the phenomena – to begin categorizing by type of company, type of manager, and type of problem. They next explored the association between these types, and the outcomes of interest. This portion of Bowen's course had an *inductive* architecture that moved up the theory pyramid. Then armed with their preliminary body of theory, Bowen and his students cycled down the deductive side of the pyramid to examine more companies in a broader range of circumstances. This allowed them to discover things that their initial theories could not explain; and to improve their constructs, refine their classification scheme, and improve their understanding of what causes what, and why.

A second circumstance is where well-researched theories pertaining to a field of management already exist. In this situation, a *deductive* course architecture can help professors improve the theories. For example, Clayton Christensen's case-based course, Building a Sustainable Enterprise, is designed deductively. For each class, students read two documents – a paper that summarizes a normative theory about a dimension of a general manager's job, and a case about a company facing a problem that is relevant to that theory. In class discussions, students then look through the lenses of the theory, to see if it accurately explains what historically happened in the company. They also use the theory to discuss what management actions will and will not lead to the desired outcomes, given the situation the company is in. Students often discover an anomaly in these complicated cases that enables the class to revisit the crispness of definitions, the categorization scheme, and the associated statement of causality. Students follow this process, theory after theory, class after class, for the semester. In the process, the theories are refined, and the students learn not just how to use theory, but how to improve it.[14]

As these experiences suggest, the dichotomy that many see between teaching and research need not create conflict. Professors who simply lecture have difficulty escaping this trade-off, of course. But for faculty who are willing follow the process of theory building as they teach, it may be better to view developing and teaching courses as *course research*. And there are two circumstances in which professors might find themselves. When a body of theory has not yet coalesced, an inductive architecture is productive. When useful theory already has emerged, then a deductive architecture can make sense. In both circumstances, however, instructors whose interest is to build theory and help students learn how to use theory, can harness the brainpower of their students by leading them through cycles up and down the theory-building pyramid.

Summary

We hope that this paper can constitute the beginnings of a pyramid of its own–a normative theory of theory building. At its foundation is the description in this paper of what we and other scholars have observed about how communities of scholars can build bodies of understanding that cumulatively improve. We have offered a set of constructs – labeled with terms such as observation, categorization, association, anomaly, descriptive theory and normative theory. Their purpose is to abstract up from the detail of thousands of research projects, to offer a general description of the way productive research processes work.

Our statements of cause and effect are, we hope, quite clear: Following the process results in good theory. Skipping a step – particularly categorization – yields flimsy theory. If researchers ask anomaly-seeking research questions and design research that maximizes the probability of finding one, it will create the opportunity for them to improve the categorization scheme and measurement methods, and to articulate better theory. Their theories will be internally valid if they examine the phenomena and outcomes of interest through the lens of all potentially relevant perspectives. And they can establish its external validity by getting the categories right. If a subsequent researcher uncovers an anomaly to a prior scholar's work, it represents triumph for both – because the prior scholar set in place a foundation that was solid enough upon which the subsequent researcher could build.

We aren't finished, of course. This theory of theory building needs a categorization scheme, and the definitions of the critical constructs must become less ambiguous. We simply hope that these concepts can provide willing researchers with a start. As scholars use these statements of cause and effect, we hope that most of them find that the statements of causality actually *do* lead to the creation of better theory. We can expect, however, that anomalies will arise, as researchers follow these methods and fail. When this happens, we invite them to publish a paper about the anomaly – focusing on what it was about the circumstance of their research that caused the causal mechanism we have described here not to work. It is only by doing so that we can approach a better understanding of what research methods can most predictably yield valuable insight–and how those methods need to vary by circumstance.

References

Allison, G. (197), *The Essence of Decision*. Glenview, IL: Scott, Foresman & Co.

Argyris, C. (1993), *On Organizational Learning*. Cambridge, MA: Blackwell.

Argyris, C. & Schon, D. (1976), *Theory in Practice*. San Francisco: Jossey-Bass.

Baldwin, C. and Clark, K.B. (2000), *Design Rules: The Power of Modularity*. Cambridge, MA: MIT Press.

Barley, S.R. (1986), "Technology as an occasion for structuring: Evidence from observations of CT scanners and the social order of radiology departments." *Administrative Science Quarterly*, 31, 1: 78-108.

Bazerman, M.H. (2005), "Conducting influential research: the need for prescriptive implications," *Academy of Management Review* 30, 1, pp. 25-31.

Black, L., Repenning, N. and Carlile, P.R. (2005), "Formalizing theoretical insights from ethnographic evidence: Revisiting Barley's study of CT-Scanning implementations." Under revision, *Administrative Science Quarterly*.

Bourdieu, P. (1989/1998), *Practical Reason*. Stanford: Stanford University Press.

Bourdieu, P. and Wacquant, L. (1992), *An Invitation to Reflexive Sociology*. Chicago: University of Chicago Press.

Bower, Joseph (1970), *Managing the Resource Allocation Process*. Englewood Cliffs, NJ: Irwin.

Bower, J.L., and Gilbert, C.G., eds. (2005), *From Resource Allocation to Strategy*. Oxford University Press.

Burgelman, Robert & Leonard Sayles (1986), *Inside Corporate Innovation*. New York: The Free Press.

Burgelman, Robert (2002), *Strategy Is Destiny*. New York: The Free Press.

Campbell, D.T.and Stanley, J.C. (1963), *Experimental and Quasi-experimental Design for Research*. Boston: Hougthon Mifflin Press.

Carlile, P.R. (2003), "Transfer, translation and transformation: Integrating approach in sharing and assessing knowledge across boundaries." Under revision, *Organization Science*.

Carr, E.H. (1961), *What Is History*? New York: Vintage Books.

Chandler, A. D. Jr. (1977), *The Visible Hand: The Managerial Revolution in American Business*. Cambridge, MA: Belknap Press.

Chandler, A. D. Jr. (1990), *Scale and Scope: The Dynamics of Industrial Capitalism*. Cambridge, MA: The Belknap Press.

Christensen, C.M. (1997), *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*. Boston: Harvard Business School Press.

Chesbrough, H.W. (1999). "The Differing organizational impact of technological change: A comparative theory of institutional factors." *Industrial and Corporate Change*, 8: 447-485.

Clegg, Brian (2003), *The First Scientist: A Life of Roger Bacon*. New York: Carroll & Graf Publishers.

Daneels, Erwin (2005), "The Effects of Disruptive Technology on Firms and Industries," *Journal of Product Innovation Management* (forthcoming special issue that focuses on this body of theory).

Ferraro, F., and J. Pfeffer & R. Sutton (2005), "Economics language and assumptions: how theories can become self-fulfilling," *Academy of Management Review* 30, 1, pp. 8-24.

Fleming, L. and Sorensen, O. (2001), "Technology as a complex adaptive system: Evidence from patent data." *Research Policy,* 30: 1019-1039.

Gilbert, C.G. (2001), *A Dilemma in Response: Examining the Newspaper Industry's Response to the Internet*. Unpublished DBA thesis, Harvard Business School.

Gilbert, C.G., and Christensen, C.M. (2005). "Anomaly Seeking Research: Thirty Years of Development in Resource Allocation Theory." In Bower, J.L., and Gilbert, C.G., eds. *From Resource Allocation to Strategy*. Oxford University Press, forthcoming.

Glaser, B. & Straus, A. (1967), *The Discovery of Grounded Theory: Strategies of Qualitative Research*. London: Wiedenfeld and Nicholson.

Hambrick, D. (1994), "1993 Presidential address: What if the Academy actually mattered?" *Academy of Management Review* 19:11-16.

Hayes, R. (1985), "Strategic Planning: Forward in Reverse?" *Harvard Business Review*, November-December: 111-119.

Hayes, R. (2002), "The History of Technology and Operations Research," Harvard Business School Working paper.

Hayes, R. and Abernathy, W. (1980), "Managing our Way to Economic Decline." *Harvard Business Review*, July-August: 7-77.

Hayes, R. and Wheelwright, S.C. (1984), *Restoring our Competitive Edge*. New York: John Wiley & Sons.

Hayes, R., Wheelwright, S. and Clark, K. (1988), *Dynamic Manufacturing*. New York: The Free Press.

Henderson, R.M. & Clark, K.B. (1990), "Architectural Innovation: The Reconfiguration of Existing Systems and the Failure of Established Firms." *Administrative Science Quarterly*, 35: 9-30.

Hunter, S.D. (2002), "Information Technology, Organizational Learning and Firm Performance." MIT/Sloan Working Paper.

Hutton, A., Miller, G., and Skinner, D. (2000), "Effective Voluntary Disclosure." Harvard Business School working paper.

James, W. (1907), *Pragmatism*. New York: The American Library.

Johnson, H.T. & Kaplan, R. (1987), *Relevance Lost*. Boston: Harvard Business School Press.

Kaplan, A. (1964), *The Conduct of Inquiry: Methodology for Behavioral Research*. Scranton, PA: Chandler.

Kaplan, R. (1986), "The role for Empirical Research in Management Accounting." *Accounting, Organizations and Society*, 4: 429-452.

Kuhn, T. (1962), *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press, 1962.

Lawrence, P. R. and Lorsch, J.W. (1967), *Organization and Environment*. Boston: Harvard Business School Press.

Leonard, D. (1995), *Wellsprings of Knowledge*. Boston: Harvard Business School Press.

Poole, M. & Van de Ven, A. (1989), "Using Paradox to Build Management and Organization Theories." *Academy of Management Review* 14: 562-578.

Popper, K. (1959), *The Logic of Scientific Discovery*. New York: Basic Books.

Porter, M. (1980), *Competitive Strategy*. New York: The Free Press.

Porter, M. (1985), *Competitive Advantage*. New York: The Free Press.

Porter, M. (1991), *The Competitive Advantage of Nations*. New York: The Free Press.

Raman, Ananth, (need citation)

Roethlisberger, F. (1977), *The Elusive Phenomena*. Boston: Harvard Business School Press.

Rumelt, Richard P. (1974), *Strategy, Structure and Economic Performance*. Cambridge, MA: Harvard University Press.

Selznick, P. (1957), *Leadership in Administration: A Sociological Interpretation*. Berkeley: University of California Press.

Simon, H. (1976), *Administrative Behavior* (3rd edition). New York: The Free Press.

Solow, R. M. (1985), "Economic History and Economics." *The American Economic Review*, 75: 328-331.

Sorensen, O. and Sorensen, J. (2001), Research Note - Finding the right mix: Franchising, organizational learning, and chain performance. *Strategic Management Journal*, 22: 713-724.

Sorensen, J. (2002), "The Strength of Corporate Culture and the Reliability of Firm Performance," *Administrative Science Quarterly*, 47: 70-91.

Spear, S.C. and Bowen, H.K. (1999), "Decoding the DNA of the Toyota production system." *Harvard Business Review*, September-October.

Stinchcombe, Arthur L. (1968), *Constructing Social Theories*." New York: Harcourt, Brace & World.

Sull, D. N. (2000), "Industrial Clusters and Organizational Inertia: An Institutional Perspective." Harvard Business School working paper.

Van de Ven, A. (2000), "Professional Science for a Professional School." In Beer, M. and Nohria, N. (Eds), *Breaking the Code of Change*. Boston: Harvard Business School Press.

Weick, K. (1989), "Theory Construction as Disciplined Imagination," *Academy of Management Review*, 14: 516-532.

Womack, J. P., Jones, D. T. & Roos, D. (1990), *The Machine that Changed the World*. New York: Rawson Associates.

Yin, R. (1984), *Case Study Research*. Beverly Hills: Sage Publications.

Yoffie, David, Sasha Mattu & Ramon Casadesus-Masanell (2002), "Intel Corporation, 1968-2003," Harvard Business School case #9-703-427.

---

[1] This model was developed first by synthesizing models of theory building that have been developed by scholars of this process in a range of fields. These include Kuhn (1962) and Popper (1959) in the natural sciences; and Campbell & Stanley (1963), A. Kaplan (1964), Glaser & Strauss (1967) Stinchcombe (1968), Roethlisberger (1977) Simon (1976), Yin (1984), R. Kaplan (1986), Weick (1989),Eisenhardt (1989) and Van de Ven (2000) in the study of management and social science. To this synthesis we have added our own observations, derived from studying various doctoral students' research efforts at Harvard, MIT, Stanford and the University of Michigan.

[2] Popper asserts that a researcher in this phase, when the theory accurately predicted what he observed, can only state that his test or experiment of the theory "corroborated" or "failed to dis-confirm" the theory.

[3] Kant, Popper, Feyerabend and others have noted that all observations are shaped, consciously or unconsciously, by cognitive structures, previous experience or some theory-in-use. While it is true that individual researchers might start their work at the top of the pyramid, generally the hypotheses that deductive theorists test have been derived consciously or unconsciously, by themselves or others, from an inductive source. We believe that few blue-sky hypotheses are conceived *de novo* at the top of the pyramid in the complete absence of observation.

[4] As we have presented preliminary versions of this paper in various faculty seminars, we have frequently found ourselves engaged in esoteric discussions about whether absolute truth exists, let alone whether we can ever discover what it is. We have concluded from these discussions that we cannot judge the value of a theory by whether it is

"true." As shown below, the best we can hope for is a body of understanding that asymptotically approaches truth. Hence, the value of a theory is assessed by its predictive power. This is why we assert that normative theory is more advanced, and more useful, than descriptive theory is.

[5] Whether or not this set can ever be defined in permanent, unambiguous ways is addressed below.

[6] Bazerman (2005) has noted that one reason why their research generally has had little influence on management is that most social science researchers choose not to be prescriptive. In fact, a culture of sorts has emerged amongst many social science researchers that descriptive theory is as far as they should go. Bazerman shows that normative theory is not only possible to develop in the social sciences, it is desirable. Ferraro, Pfeffer and Sutton (2005) seem to agree that normative or prescriptive social science theories *can* profoundly influence behavior – sometimes in self-fulfilling ways.

[7] Need reference for Hackman's article on cockpit teams.

[8] The observation that management research is often faddish has been raised enough that it no longer seems shocking (Micklethwait and Wooldridge, 1996; Abrahamson, 1998). It need not be faddish, however, if researchers would design and report their research in ways that reflect the importance of getting the categories right.

[9] In general, the discovery of anomalies in normative theory sends researchers back to the phenomena and categories of the pyramid of normative theory. Kuhn noted, however, that on occasion the accumulated weight of unresolvable anomalies will take down a paradigm. This can throw the theory-building enterprise back to the descriptive theory stage, as a preliminary stage to the emergence of a new paradigm.

[10] Personal conversations between Clayton Christensen and Mr. Teo Min Kian, Chairman of the Economic Development Board, government of Singapore, 2003-2005.

[11] Spear is now engaged in search of anomalies on the deductive side of the cycle of building normative theory. Because no company besides Toyota has employed this causal mechanism, Spear cannot retrospectively study other companies. Like Johnson & Kaplan did when they used "action research" to study the implementation problems of activity-based costing, Spear is helping companies in very different circumstances to use his statements of causality, to see whether the mechanism of these four rules yields the same results. To date, companies in industries as diverse as aluminum smelting, hospitals, and jet engine design have achieved the results that Spear's theory predicts – he has not yet succeeded in finding an anomaly. The categorization step of this body of normative theory still has no salient boundaries within it.

[12] Some large sample, quantitative studies in strategy research have begun to turn to analyses that estimate simultaneously the expected value (a mean analysis) and the variance associated with performance oriented dependent variables using a "variance decomposition" approach (Fleming and Sorensen, 2001; Sorensen and Sorensen, 2001). The simultaneous nature of this methodological approach allows a deeper understanding of the mean as well as the variance associated with a firm overtime (Sorensen, 2002) or a population of firms (Hunter, 2002). What such analysis suggests is that when there is significant heterogeneity in a given strategic environment, not only will there be variance in firm performance, but also what a firm needs to do to be successful will also differ based of the niche that they pursue. This reminds us that explanations for strategic questions are not only contingent, but more importantly are based on an understanding what sources of variance, what relations across different variables, matter most and why. From a methodological point of view, this also reminds of how our abilities (i.e., tools, methods) to represent data shape how we are able to describe what "strategic action" is possible.

[13] In this light, the recent decision by the editors of the *Academy of Management Review* to require all papers that use field-based data to go through three mandatory revise-and-resubmit reviews, whereas those using quantitative data do not, represents an extraordinarily naïve view of where numeric data come from. They shame the Academy with this hypocritical policy.

14 At one point Christensen attempted to teach his course through an inductive architecture. Case by case, he attempted to lead his students to discover well-documented theories that prior scholars already had discovered. The course was a disaster – the wrong architecture for the circumstance. Students could tell that Christensen already had the answer, and his attempts to orchestrate a case discussion seemed like the professor was asking the students to guess what was on his mind. The next year, Christensen revised his course to the deductive architecture described above, and students reacted very positively to the same material.